Agilent Feature Extraction Software (v9.5)

Reference Guide

Research Use Only

Agilent Technologies

Notices

Edition

Fourth Edition, February 2007

Printed in USA

Agilent Technologies, Inc.
5301 Stevens Creek Blvd.
Santa Clara, CA 95051

Agilent Recognized Trademarks

Adobe®; Acrobat®–Adobe, the Adobe Logo, Acrobat and the Acrobat Logo are trademarks of Adobe Systems Incorporated.

Pentium®–Pentium is a U.S. registered trademark of Intel Corporation.

Microsoft®–Microsoft is a U.S. registered trademark of Microsoft Corporation.

Rosetta Luminator™–Rosetta Luminator is a trademark of Rosetta Inpharmatics LLC.

Rosetta Resolver®–Rosetta Resolver is a U.S. registered trademark of Rosetta Inpharmatics LLC.

Windows NT®–Windows NT is a U.S. registered trademark of Microsoft Corporation.

Windows® and MS Windows®–Windows and MS Windows are U.S. registered trademarks of Microsoft Corporation.

Patents

Portions of this product may be covered under US patent 6571005 licensed from the Regents of the University of California.

Warranty

The material contained in this document is provided "as is," and is subject to being changed, without notice, in future editions. Further, to the maximum extent permitted by applicable law, Agilent disclaims all warranties, either express or implied, with regard to this manual and any information contained herein, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Agilent shall not be liable for errors or for incidental or consequential damages in connection with the furnishing, use, or performance of this document or of any information contained herein. Should Agilent and the user have a separate written agreement with warranty terms covering the material in this document that conflict with these terms, the warranty terms in the separate agreement shall control.

Technology Licenses

The hardware and/or software described in this document are furnished under a license and may be used or copied only in accordance with the terms of such license.

Restricted Rights Legend

U.S. Government Restricted Rights. Software and technical data rights granted to the federal government include only those rights customarily provided to end user customers. Agilent provides this customary commercial license in Software and technical data pursuant to FAR 12.211 (Technical Data) and 12.212 (Computer Software) and, for the Department of Defense, DFARS 252.227-7015 (Technical Data - Commercial Items) and DFARS 227.7202-3 (Rights in Commercial Computer Software or Computer Software Documentation).

Safety Notices

CAUTION

A **CAUTION** notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in damage to the product or loss of important data. Do not proceed beyond a **CAUTION** notice until the indicated conditions are fully understood and met.

WARNING

A WARNING notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in personal injury or death. Do not proceed beyond a WARNING notice until the indicated conditions are fully understood and met.

In This Guide...

This *Reference Guide* contains tables that list default parameter values and results for Feature Extraction (FE) analyses and explanations of how FE uses its algorithms to calculate results.

1 Protocol Default Settings

This chapter includes tables that list the default parameter values found in the protocols shipped with the software (2-color gene expression (GE), 1-color GE, CGH and non-Agilent protocols).

2 QC Report Results

Learn how to read and interpret the QC Reports.

3 Text File Parameters and Results

This chapter contains a listing of parameters and results within the text file produced after Feature Extraction.

4 XML (MAGE-ML) Results

Refer to this chapter to find the results contained in the MAGE-ML files generated after Feature Extraction.

5 How Algorithms Calculate Results

Learn how Feature Extraction algorithms calculate the results that help you interpret your gene expression (2-color and 1-color) and CGH experiments.

6 Command Line Feature Extraction

This chapter contains the commands and arguments to integrate Feature Extraction into a completely automated workflow.

Acknowledgments

Apache acknowledgment

Part of this software is based on the Xerces XML parser, Copyright (c) 1999-2000 The Apache Software Foundation. All Rights Reserved (www.apache.org).

JPEG acknowledgment

This software is based in part on the work of the Independent JPEG Group. Copyright (c) 1991-1998, Thomas G. Lane. All Rights Reserved.

Loess/Netlib acknowledgment

Part of this software is based on a Loess/Lowess algorithm and implementation. The authors of Loess/Lowess are Cleveland, Grosse and Shyu. Copyright (c) 1989, 1992 by AT&T. Permission to use, copy, modify and distribute this software for any purpose without fee is hereby granted, provided that this entire notice in included in all copies of any software which is or includes a copy or modification of this software and in all copies of the supporting documentation for such software.

THIS SOFTWARE IS BEING PROVIDED “AS IS”, WITHOUT ANY EXPRESS OR IMPLIED WARRANTY. NEITHER THE AUTHORS NOR AT&T MAKE ANY REPRESENTATION OR WARRANTY OF ANY KIND CONCERNING THE MERCHANTABILITY OF THIS SOFTWARE OR ITS FITNESS FOR ANY PARTICULAR PURPOSE.

Stanford University School of Medicine acknowledgment

Non-Agilent microarray image courtesy of Dr. Roger Wagner, Division of Cardiovascular Medicine, Stanford University School of Medicine

Ultimate Grid acknowledgment

This software contains material that is Copyright (c) 1994-1999 DUNDAS SOFTWARE LTD., All Rights Reserved.

LibTiff acknowledgement

Part of this software is based upon LibTIFF version 3.8.0.

Contents

1 Default Protocol Settings

When the software assigns a protocol to an extraction set, the software loads a set of protocol parameter values and settings that affect the process and results for Feature Extraction.

See Chapter 4, "Changing Protocol Settings" in the User Guide to learn the purpose of all the parameters and settings and how to modify them.

Parameter values in the protocol depend on the microarray type and your experiment. The following pages list the default settings for each of the protocol templates shipped with the software. Each protocol template represents a different microarray type. You can view these settings and values when you open the Protocol Editor for each of the protocol templates.

Agilent protocols are meant for use with Agilent microarrays scanned with either an Agilent or GenePix (4000A or 4000B) scanner. The non-Agilent protocol is meant for use with non-Agilent microarrays that are scanned with an Agilent scanner.

Default protocol settings—an introduction

To learn more about changing the default values for the protocols, see "View or change the protocol properties" on page 146 of the User Guide.

To learn about the naming of the protocol templates, see "Protocol templates" on page 142 of the User Guide.

Agilent also provides new and updated protocol templates. To download these protocol templates, go to http://www.agilent.com/chem/feprotocols and follow the instructions for importing the templates into the Protocol Browser.

This chapter presents tables for viewing the default settings for each protocol. Parameter values depend on:

- microarray type
- lab protocol
- formats
- scanner used

Listed below are the names of the non-removable protocol templates and where you can find the tables listing their default values in this chapter.

Table 1 Location of protocol template default settings

Protocol Template Name	Location in chapter
GE1-v1_95_Feb07	page 14
GE1-v5_95_Feb07	page 18
GE2-SSPE_95_Feb07	page 22
GE2-v4_95_Feb07	page 27
GE2-v5_95_Feb07	page 32
CGH-v4_95_Feb07	page 37
GE2-nonAT_95_Feb07	page 42

Difference between CGH and gene expression microarrays

To see the differences in some default settings between protocols, go to "Differences in protocol settings based on each step" on page 45.

CGH microarrays possess a different negative control sequence scheme than the gene expression microarrays. The gene expression microarrays have many replicate negative control features using only one sequence. The CGH microarrays have many sequences of negative controls that span the range of sequence variability seen in the biological probes used on the microarrays. This difference in the control grid (especially the multiple sequences used for negative controls) leads to a difference in protocol settings.

Hidden Settings

To create a protocol for a specific type of microarray, you ***must*** use a protocol template or existing protocol for the same type of microarray.

CAUTION Protocol templates provide both visible and **hidden** settings whose values are specific to the type or format of microarrays. Although you can change the visible settings so that any two protocols of different type *appear* identical, you **cannot change the hidden settings** that distinguish these protocols from one another.

One of the hidden settings is the specification of the QC report. A 2-color protocol generates a 2-color QC report; a 1-color protocol generates a 1-color QC report; and a CGH protocol generates a CGH QC report.

The tables show only the default visible parameter values for the steps of the protocol. You can view the hidden parameters in the FE PARAMS table. See "Parameters/options (FEPARAMS)" on page 113. Many of these hidden parameters are image processing ones that will be chosen using the "Automatically Determine" function.

Tables of Default Protocol Settings

GE1-v1_95 protocol

This is a 1-color gene expression protocol for use with the Gene Expression Analysis lab protocol (version 1-publication number G4140-90040).

Table 2 Default settings for GE1-v1_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
Place Grid	Array Format	For any format automatically determined or selected by you, the software uses the default Placement Method listed below.	Automatically Determine [Recognized formats: Single Density (11k, 22k), 25k, Double Density (44k), 95k, 185k (5 and 10 u), 244k (5 and 10 u) and Third Party]
	Placement Method	The parameters and values for placing the grid differ depending on the format, but you can't see the differences because the values are hidden.	Allow Some Distortion
Find Spots	Spot Format	Depending on the format selected by the software or by you, the default settings for this step change. See the rows below for the default values for finding spots.	Automatically Determine [Recognized formats: same as those listed above]
		Use the Nominal Diameter from the Grid Template	True (All Formats)
		Spot Deviation Limit	1.50 (All Formats)
		Calculation of Spot Statistics Method	Use Cookie (All Formats)
		Cookie Percentage	0.650 (Single Density, 25k)
			0.561 (Double Density, 95k)
			0.700 (185k, 244k)

Table 2 Default settings for GE1-v1_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Exclusion Zone Percentage	1.200 (All Formats)
		Auto Estimate the Local Radius	True (All except 185, 244)
			False (185k, 244k)
		LocalBGRadius	100 (when False for 185k, 244k)
	Pixel Outlier Rejection Method		Inter Quartile Region (Automatically Determine and All Formats)
		RejectIQRFeat	1.42 (All Formats)
		RejectIQRBG	1.42 (All Formats)
	Statistical Method for Spot Values from Pixels		Use Mean/Standard Deviation (Automatically Determine and All Formats)
Flag Outliers	Compute Population Outliers		True
		Minimum Population	10
		IQRatio	1.42
		Background IQRatio	1.42
		Use Qtest for Small Populations?	False
	Compute Non Uniform Outliers		True
	Scanner	The values for the parameters change depending on the scanner used for the image. See below for differences.	Automatically Determine
	Agilent scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.01690
		Green Poissonian Noise Term Multiplier	40
		Green Signal Constant Term Multiplier	1

Table 2 Default settings for GE1-v1_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Background – (%CV)^2	0.09000
		Green Poissonian Noise Term Multiplier	3
		Green Background Constant Term Multiplier	1
	GenePix (Axon) scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.01440
		Green Poissonian Noise Term Multiplier	20
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Green Poissonian Noise Term Multiplier	1
		Green Background Constant Term Multiplier	1
Compute Bkgd, Bias and Error	Background Subtraction Method		No Background Subtraction
	Significance (for IsPosAndSignif and IsWellAboveBG)		Use Pixel Statistics for Significance
		2-sided t-test of feature vs background max p-value	0.01
		WellAboveMulti	2.6
	Signal Correction—Calculate Surface Fit (required for Spatial Detrend)		True
		Feature Set for Surface Fit	FeaturesInNegativeControlRange
		Perform Filtering for Surface Fit	False
		Perform Spatial Detrending	True

Table 2 Default settings for GE1-v1_95 protocol

Protocol Step	Parameter	Default Setting/Value (v.9.5)
	Signal Correction—Adjust Background Globally	False
	Signal Correction—Perform Multiplicative Detrending	True
	Detrend on Replicates Only	False
	Filter Low signal probes from Fit?	True
	Neg. Ctrl. Threshold Mult. Detrend Factor	3
	Perform Filtering for Fit	Use Window Average
	Use polynomial data fit instead of LOESS?	True
	Polynomial Multiplicative DetrendDegree	2
	Robust Neg Ctrl Stats?	False
	Choose universal error, or most conservative	Most Conservative
	MultErrorGreen	0.1000
	Auto Estimate Add Error Green	True
	Use Surrogates	False
Calculate Metrics	Spikein Target Used	True
	Min Population for Replicate Stats?	5
	PValue for Differential Expression	0.010000
	Percentile Value	75
Generate Results	Generate Single Text File	True
	JPEG Down Sample Factor	4

GE1-v5_95 protocol

This is a 1-color gene expression protocol for use with the Gene Expression Analysis lab protocol (version 5 -publication number G4140-90040 V.5.5).

Table 3 Default settings for GE1-v5_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
Place Grid	Array Format	For any format automatically determined or selected by you, the software uses the default Placement Method listed below.	Automatically Determine [Recognized formats: Single Density (11k, 22k), 25k, Double Density (44k), 95k, 185k (5 and 10 u), 244k (5 and 10 u) and Third Party]
	Placement Method	The parameters and values for placing the grid differ depending on the format, but you can't see the differences because the values are hidden.	Allow Some Distortion
Find Spots	Spot Format	Depending on the format selected by the software or by you, the default settings for this step change. See the rows below for the default values for finding spots.	Automatically Determine [Recognized formats: same as those listed above]
		Use the Nominal Diameter from the Grid Template	True (All Formats)
		Spot Deviation Limit	1.50 (All Formats)
		Calculation of Spot Statistics Method	Use Cookie (All Formats)
		Cookie Percentage	0.650 (Single Density, 25k)
			0.561 (Double Density, 95k)
			0.700 (185k, 244k)
		Exclusion Zone Percentage	1.200 (All Formats)
		Auto Estimate the Local Radius	True (All except 185, 244)
			False (185k, 244k)

Table 3 Default settings for GE1-v5_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		LocalBGRadius	100 (when False for 185k, 244k)
	Pixel Outlier Rejection Method		Inter Quartile Region (Automatically Determine and All Formats)
		RejectIQRFeat	1.42 (All Formats)
		RejectIQRBG	1.42 (All Formats)
	Statistical Method for Spot Values from Pixels		Use Mean/Standard Deviation (Automatically Determine and All Formats)
Flag Outliers	Compute Population Outliers		True
		Minimum Population	10
		IQRatio	1.42
		Background IQRatio	1.42
		Use Qtest for Small Populations?	True
	Compute Non Uniform Outliers		True
	Scanner	The values for the parameters change depending on the scanner used for the image. See below for differences.	Automatically Determine
	Agilent scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.04
		Green Poissonian Noise Term Multiplier	20
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Green Poissonian Noise Term Multiplier	3

Table 3 Default settings for GE1-v5_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Green Background Constant Term Multiplier	1
	GenePix (Axon) scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.01440
		Green Poissonian Noise Term Multiplier	30
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Green Poissonian Noise Term Multiplier	1
		Green Background Constant Term Multiplier	1
Compute Bkgd, Bias and Error	Background Subtraction Method		No Background Subtraction
	Significance (for IsPosAndSignif and IsWellAboveBG)		Use Error Model for Significance
		2-sided t-test of feature vs background max p-value	0.01
		WellAboveMulti	13
	Signal Correction—Calculate Surface Fit (required for Spatial Detrend)		True
		Feature Set for Surface Fit	FeaturesInNegativeControlRange
		Perform Filtering for Surface Fit	True
		Perform Spatial Detrending	True
	Signal Correction—Adjust Background Globally		False
	Signal Correction—Perform Multiplicative Detrending		True
		Detrend on Replicates Only	True

Table 3 Default settings for GE1-v5_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Filter Low signal probes from Fit?	True
		Neg. Ctrl. Threshold Mult. Detrend Factor	5
		Perform Filtering for Fit	User Window Average
		Use polynomial data fit instead of LOESS?	True
		Polynomial Multiplicative DetrendDegree	4
	Robust Neg Ctrl Stats?		False
	Choose universal error, or most conservative		Most Conservative
		MultErrorGreen	0.1000
		Auto Estimate Add Error Green	True
	Use Surrogates		True
Calculate Metrics	Spikein Target Used		True
	Min Population for Replicate Stats?		5
	PValue for Differential Expression		0.010000
	Percentile Value		75
Generate Results	Generate Single Text File		True
	JPEG Down Sample Factor		4

GE2-SSPE_95 protocol

This is a 2-color gene expression protocol for use with the previous (SSPE or SSC) versions of the lab protocol.

Table 4 Default settings for GE2-SSPE_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
Place Grid	Array Format	For any format automatically determined or selected by you, the software uses the default Placement Method listed below.	Automatically Determine [Recognized formats: Single Density (11k, 22k), 25k, Double Density (44k), 95k, 185k (5 and 10 u), 244k (5 and 10 u) and Third Party]
	Placement Method	The parameters and values for placing the grid differ depending on the format, but you can't see the differences because the values are hidden.	Allow Some Distortion
Find Spots	Spot Format	Depending on the format selected by the software or by you, the default settings for this step change. See the rows below for the default values for finding spots.	Automatically Determine [Recognized formats: same as those listed above]
		Use the Nominal Diameter from the Grid Template	True (All Formats)
		Spot Deviation Limit	1.50 (All Formats)
		Calculation of Spot Statistics Method	Use Cookie (All Formats)
		Cookie Percentage	0.650 (Single Density, 25k)
			0.561 (Double Density, 95k)
			0.700 (185k, 244k)
		Exclusion Zone Percentage	1.200 (All Formats)
		Auto Estimate the Local Radius	True (All except 185, 244)
			False (185k, 244k)

Table 4 Default settings for GE2-SSPE_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		LocalBGRadius	100 (when False for 185k, 244k)
	Pixel Outlier Rejection Method		Inter Quartile Region (Automatically Determine and All Formats)
		RejectIQRFeat	1.42 (All Formats)
		RejectIQRBG	1.42 (All Formats)
	Statistical Method for Spot Values from Pixels		Use Mean/Standard Deviation (Automatically Determine and All Formats)
Flag Outliers	Compute Population Outliers		True
		Minimum Population	10
		IQRatio	1.42
		Background IQRatio	1.42
		Use Qtest for Small Populations?	False
	Compute Non Uniform Outliers		True
	Scanner	The values for the parameters change depending on the scanner used for the image. See below for differences.	Automatically Determine
	Agilent scanner		
	Automatically Compute OL Polynomial Terms		False
		Feature –(%CV)^2	0.00810
		Poissonian Noise Term	320 (R, G combined)
		Background Term	600 (R, G combined)
		Background –(%CV)^2	0.2250
		Poissonian Noise Term	320 (R, G combined)
		Background Term	600 (R, G combined)

Table 4 Default settings for GE2-SSPE_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
	GenePix (Axon) scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.01440
		Red Poissonian Noise Term Multiplier	30
		Red Signal Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	20
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Red Poissonian Noise Term Multiplier	1
		Red Background Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	1
		Green Background Constant Term Multiplier	1
Compute Bkgd, Bias and Error	Background Subtraction Method		No Background Subtraction
	Significance (for IsPosAndSignif and IsWellAboveBG)		Use Pixel Statistics for Significance
		2-sided t-test of feature vs background max p-value	0.01
		WellAboveMulti	2.6
	Signal Correction—Calculate Surface Fit (required for Spatial Detrend)		True
		Feature Set for Surface Fit	AllFeatureTypes

Table 4 Default settings for GE2-SSPE_95 protocol

Protocol Step	Parameter	Default Setting/Value (v.9.5)
	Perform Filtering for Surface Fit	True
	Perform Spatial Detrending	True
	Signal Correction—Adjust Background Globally	False
	Signal Correction—Perform Multiplicative Detrending	False
	Robust Neg Ctrl Stats?	False
	Choose universal error, or most conservative	Most Conservative
	MultErrorGreen	0.0800
	MultErrorRed	0.0800
	Auto Estimate Add Error Red	True
	Auto Estimate Add Error Green	True
	Use Surrogates	True
Correct Dye Biases	Dye Normalization Probe Selection Method	Use Rank Consistent Probes
	Rank Tolerance	0.050
	Variable Rank Tolerance	False
	Max Number Ranked Probes	-1
	Omit Background Population Outliers	False
	Allow Positive and Negative Controls	False
	Signal Characteristics	OnlyPositiveAndSignificantSignals
	Normalization Correction Method	Linear and Lowess
Compute Ratios	Peg Log Ratio Value	4.00
Calculate Metrics	Spikein Target Used	True
	Min Population for Replicate Stats?	5
	PValue for Differential Expression	0.010000
	Percentile Value	75

Table 4 Default settings for GE2-SSPE_95 protocol

Protocol Step	Parameter	Default Setting/Value (v.9.5)
Generate Results	Generate Single Text File	True
	JPEG Down Sample Factor	4

GE2-v4_95 protocol

This is a 2-color gene expression protocol for use with the 2-color Microarray-based Gene Expression Analysis lab protocol (version 4–publication number G4140-90050) .

Table 5 Default settings for GE2-v4_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
Place Grid	Array Format	For any format automatically determined or selected by you, the software uses the default Placement Method listed below.	Automatically Determine [Recognized formats: Single Density (11k, 22k), 25k, Double Density (44k), 95k, 185k (5 and 10 u), 244k (5 and 10 u) and Third Party]
	Placement Method	The parameters and values for placing the grid differ depending on the format, but you can't see the differences because the values are hidden.	Allow Some Distortion
Find Spots	Spot Format	Depending on the format selected by the software or by you, the default settings for this step change. See the rows below for the default values for finding spots.	Automatically Determine [Recognized formats: same as those listed above]
		Use the Nominal Diameter from the Grid Template	True (All Formats)
		Spot Deviation Limit	1.50 (All Formats)
		Calculation of Spot Statistics Method	Use Cookie (All Formats)
		Cookie Percentage	0.650 (Single Density, 25k)
			0.561 (Double Density, 95k)
			0.700 (185k, 244k)
		Exclusion Zone Percentage	1.200 (All Formats)
		Auto Estimate the Local Radius	True (All except 185, 244)
			False (185k, 244k)

Table 5 Default settings for GE2-v4_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		LocalBGRadius	100 (when False for 185k, 244k)
	Pixel Outlier Rejection Method		Inter Quartile Region (Automatically Determine and All Formats)
		RejectIQRFeat	1.42 (All Formats)
		RejectIQRBG	1.42 (All Formats)
	Statistical Method for Spot Values from Pixels		Use Mean/Standard Deviation (Automatically Determine and All Formats)
Flag Outliers	Compute Population Outliers		True
		Minimum Population	10
		IQRatio	1.42
		Background IQRatio	1.42
		Use Qtest for Small Populations?	False
	Compute Non Uniform Outliers		True
	Scanner	The values for the parameters change depending on the scanner used for the image. See below for differences.	Automatically Determine
	Agilent scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.01690
		Red Poissonian Noise Term Multiplier	40
		Red Signal Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	40

Table 5 Default settings for GE2-v4_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Red Poissonian Noise Term Multiplier	3
		Red Background Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	3
		Green Background Constant Term Multiplier	1
	GenePix (Axon) scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.01440
		Red Poissonian Noise Term Multiplier	30
		Red Signal Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	20
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Red Poissonian Noise Term Multiplier	1
		Red Background Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	1

Table 5 Default settings for GE2-v4_95 protocol

Protocol Step	Parameter	Default Setting/Value (v.9.5)
	Green Background Constant Term Multiplier	1
Compute Bkgd, Bias and Error	Background Subtraction Method	No Background Subtraction
	Significance (for IsPosAndSignif and IsWellAboveBG)	Use Pixel Statistics for Significance
	2-sided t-test of feature vs background max p-value	0.01
	WellAboveMulti	2.6
	Signal Correction—Calculate Surface Fit (required for Spatial Detrend)	True
	Feature Set for Surface Fit	FeaturesInNegativeControlRange
	Perform Filtering for Surface Fit	False
	Perform Spatial Detrending	True
	Signal Correction—Adjust Background Globally	False
	Signal Correction—Perform Multiplicative Detrending	False
	Robust Neg Ctrl Stats?	False
	Choose universal error, or most conservative	Most Conservative
	MultErrorGreen	0.1000
	MultErrorRed	0.1000
	Auto Estimate Add Error Red	True
	Auto Estimate Add Error Green	True
	Use Surrogates	True
Correct Dye Biases	Dye Normalization Probe Selection Method	Use Rank Consistent Probes
	Rank Tolerance	0.050
	Variable Rank Tolerance	False
	Max Number Ranked Probes	-1
	Omit Background Population Outliers	False

Table 5 Default settings for GE2-v4_95 protocol

Protocol Step	Parameter	Default Setting/Value (v.9.5)
	Allow Positive and Negative Controls	False
	Signal Characteristics	OnlyPositiveAndSignificantSignals
	Normalization Correction Method	Linear and Lowess
Compute Ratios	Peg Log Ratio Value	4.00
Calculate Metrics	Spikein Target Used	True
	Min Population for Replicate Stats?	5
	PValue for Differential Expression	0.010000
	Percentile Value	75
Generate Results	Generate Single Text File	True
	JPEG Down Sample Factor	4

GE2-v5_95 protocol

This is a 2-color gene expression protocol for use with the 2-color Microarray-based Gene Expression Analysis lab protocol (version 5–publication number G4140-90040 V.5.5) .

Table 6 Default settings for GE2-v5_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
Place Grid	Array Format	For any format automatically determined or selected by you, the software uses the default Placement Method listed below.	Automatically Determine [Recognized formats: Single Density (11k, 22k), 25k, Double Density (44k), 95k, 185k (5 and 10 u), 244k (5 and 10 u) and Third Party]
	Placement Method	The parameters and values for placing the grid differ depending on the format, but you can't see the differences because the values are hidden.	Allow Some Distortion
Find Spots	Spot Format	Depending on the format selected by the software or by you, the default settings for this step change. See the rows below for the default values for finding spots.	Automatically Determine [Recognized formats: same as those listed above]
		Use the Nominal Diameter from the Grid Template	True (All Formats)
		Spot Deviation Limit	1.50 (All Formats)
		Calculation of Spot Statistics Method	Use Cookie (All Formats)
		Cookie Percentage	0.650 (Single Density, 25k)
			0.561 (Double Density, 95k)
			0.700 (185k, 244k)
		Exclusion Zone Percentage	1.200 (All Formats)
		Auto Estimate the Local Radius	True (All except 185, 244)
			False (185k, 244k)

Table 6 Default settings for GE2-v5_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		LocalBGRadius	100 (when False for 185k, 244k)
	Pixel Outlier Rejection Method		Inter Quartile Region (Automatically Determine and All Formats)
		RejectIQRFeat	1.42 (All Formats)
		RejectIQRBG	1.42 (All Formats)
	Statistical Method for Spot Values from Pixels		Use Mean/Standard Deviation (Automatically Determine and All Formats)
Flag Outliers	Compute Population Outliers		True
		Minimum Population	10
		IQRatio	1.42
		Background IQRatio	1.42
		Use Qtest for Small Populations?	True
	Compute Non Uniform Outliers		True
	Scanner	The values for the parameters change depending on the scanner used for the image. See below for differences.	Automatically Determine
	Agilent scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.04
		Red Poissonian Noise Term Multiplier	20
		Red Signal Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	20

Table 6 Default settings for GE2-v5_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Red Poissonian Noise Term Multiplier	3
		Red Background Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	3
		Green Background Constant Term Multiplier	1
	GenePix (Axon) scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.01440
		Red Poissonian Noise Term Multiplier	30
		Red Signal Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	20
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Red Poissonian Noise Term Multiplier	1
		Red Background Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	1

Table 6 Default settings for GE2-v5_95 protocol

Protocol Step	Parameter	Default Setting/Value (v.9.5)
	Green Background Constant Term Multiplier	1
Compute Bkgd, Bias and Error	Background Subtraction Method	No Background Subtraction
	Significance (for IsPosAndSignif and IsWellAboveBG)	Use Error Model for Significance
	2-sided t-test of feature vs background max p-value	0.01
	WellAboveMulti	13
	Signal Correction—Calculate Surface Fit (required for Spatial Detrend)	True
	Feature Set for Surface Fit	FeaturesInNegativeControlRange
	Perform Filtering for Surface Fit	True
	Perform Spatial Detrending	True
	Signal Correction—Adjust Background Globally	False
	Signal Correction—Perform Multiplicative Detrending	True
	Detrend on Replicates Only	True
	Filter Low signal probes from Fit?	True
	Neg. Ctrl. Threshold Mult. Detrend Factor	5
	Perform Filtering for Fit	Use Window Average
	Robust Neg Ctrl Stats?	False
	Choose universal error, or most conservative	Most Conservative
	MultErrorGreen	0.1000
	MultErrorRed	0.1000
	Auto Estimate Add Error Red	True
	Auto Estimate Add Error Green	True
	Use Surrogates	True

Table 6 Default settings for GE2-v5_95 protocol

Protocol Step	Parameter	Default Setting/Value (v.9.5)
Correct Dye Biases	Dye Normalization Probe Selection Method	Use Rank Consistent Probes
	Rank Tolerance	0.050
	Variable Rank Tolerance	False
	Max Number Ranked Probes	8000
	Omit Background Population Outliers	False
	Allow Positive and Negative Controls	False
	Signal Characteristics	OnlyPositiveAndSignificantSignals
	Normalization Correction Method	Linear and Lowess
Compute Ratios	Peg Log Ratio Value	4.00
Calculate Metrics	Spikein Target Used	True
	Min Population for Replicate Stats?	5
	PValue for Differential Expression	0.010000
	Percentile Value	75
Generate Results	Generate Single Text File	True
	JPEG Down Sample Factor	4

CGH-v4_95

This is a CGH protocol for use with the Oligonucleotide Array-Based CGH for Genomic DNA Analysis lab protocol (version 4 - publication number G4410-90010).

Table 7 Default settings for CGH-v4_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
Place Grid	Array Format	For any format automatically determined or selected by you, the software uses the default Placement Method listed below.	Automatically Determine [Recognized formats: Single Density (11k, 22k), 25k, Double Density (44k), 95k, 185k (5 and 10 u), 244k (5 and 10 u) and Third Party]
	Placement Method	The parameters and values for placing the grid differ depending on the format, but you can't see the differences because the values are hidden.	Allow Some Distortion
Find Spots	Spot Format	Depending on the format selected by the software or by you, the default settings for this step change. See the rows below for the default values for finding spots.	Automatically Determine [Recognized formats: same as those listed above]
		Use the Nominal Diameter from the Grid Template	True (All Formats)
		Spot Deviation Limit	1.50 (All Formats)
		Calculation of Spot Statistics Method	Use Cookie (All Formats)
		Cookie Percentage	0.650 (Single Density, 25k)
			0.561 (Double Density, 95k)
			0.700 (185k, 244k)
		Exclusion Zone Percentage	1.200 (All Formats)
		Auto Estimate the Local Radius	True (All except 185, 244)
			False (185k, 244k)

Table 7 Default settings for CGH-v4_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		LocalBGRadius	100 (when False for 185k, 244k)
	Pixel Outlier Rejection Method		Inter Quartile Region (Automatically Determine and All Formats)
		RejectIQRFeat	1.42 (All Formats)
		RejectIQRBG	1.42 (All Formats)
	Statistical Method for Spot Values from Pixels		Use Mean/Standard Deviation (Automatically Determine and All Formats)
Flag Outliers	Compute Population Outliers		True
		Minimum Population	10
		IQRatio	1.42
		Background IQRatio	1.42
		Use Qtest for Small Populations?	True
	Compute Non Uniform Outliers		True
	Scanner	The values for the parameters change depending on the scanner used for the image. See below for differences.	Automatically Determine
	Agilent scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.04
		Red Poissonian Noise Term Multiplier	5
		Red Signal Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	5

Table 7 Default settings for CGH-v4_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Green Signal Constant Term Multiplier	1
		Background – (%CV)^2	0.09000
		Red Poissonian Noise Term Multiplier	3
		Red Background Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	3
		Green Background Constant Term Multiplier	1
	GenePix (Axon) scanner		
	Automatically Compute OL Polynomial Terms		True
		Feature – (%CV)^2	0.01440
		Red Poissonian Noise Term Multiplier	30
		Red Signal Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	15
		Green Signal Constant Term Multiplier	1
		Background – %CV^2	0.09000
		Red Poissonian Noise Term Multiplier	1
		Red Background Constant Term Multiplier	1
		Green Poissonian Noise Term Multiplier	1

Table 7 Default settings for CGH-v4_95 protocol

Protocol Step	Parameter	Default Setting/Value (v.9.5)
	Green Background Constant Term Multiplier	1
Compute Bkgd, Bias and Error	Background Subtraction Method	No Background Subtraction
	Significance (for IsPosAndSignif and IsWellAboveBG)	Use Error Model for Significance
	2-sided t-test of feature vs background max p-value	0.01
	WellAboveMulti	2.6
	Signal Correction—Calculate Surface Fit (required for Spatial Detrend)	True
	Feature Set for Surface Fit	OnlyNegativeControlFeatures
	Perform Filtering for Surface Fit	False
	Perform Spatial Detrending	True
	Signal Correction—Adjust Background Globally	False
	Signal Correction—Perform Multiplicative Detrending	True
	Detrend on Replicates Only	False
	Filter Low signal probes from Fit?	True
	Neg. Ctrl. Threshold Mult. Detrend Factor	3
	Perform Filtering for Fit	User Window Average
	Use polynomial data fit instead of LOESS?	True
	Polynomial Multiplicative DetrendDegree	4
	Robust Neg Ctrl Stats?	True
	Choose universal error, or most conservative	Most Conservative
	MultErrorGreen	0.1000
	MultErrorRed	0.1000

Table 7 Default settings for CGH-v4_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Auto Estimate Add Error Red	True
		Auto Estimate Add Error Green	True
	Use Surrogates		True
Correct Dye Biases	Dye Normalization Probe Selection Method		Use Rank Consistent Probes
		Rank Tolerance	0.050
		Variable Rank Tolerance	False
		Max Number Ranked Probes	-1
	Omit Background Population Outliers		False
	Allow Positive and Negative Controls		False
	Signal Characteristics		OnlyPositiveAndSignificantSignals
	Normalization Correction Method		Linear
Compute Ratios	Peg Log Ratio Value		4.00
Calculate Metrics	Spikein Target Used		True
	Min Population for Replicate Stats?		3
	PValue for Differential Expression		0.010000
	Percentile Value		75
Generate Results	Generate Single Text File		True
	JPEG Down Sample Factor		4

GE2-nonAT_95 protocol

Use this protocol for running Feature Extraction on non-Agilent microarrays scanned with an Agilent scanner.

CAUTION These protocol settings may not be optimum for non-Agilent microarrays or Agilent microarrays processed with non-Agilent procedures. You must determine the settings and values that are optimum for your system.

Table 8 Default settings for GE2-nonAT_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
Place Grid	Array Format		Third Party
		Placement Method	Allow Some Distortion
Find Spots	Spot Format		Third Party
		Use the Nominal Diameter from the Grid Template	True
		Spot Deviation Limit	1.50
		Calculation of Spot Statistics Method	Use Cookie
		Cookie Percentage	0.650
		Exclusion Zone Percentage	1.200
		Auto Estimate the Local Radius	True
	Pixel Outlier Rejection Method		Inter Quartile Region
		RejectIQRFeat	1.42
		RejectIQRBG	1.42
	Statistical Method for Spot Values from Pixels		Use Mean/Standard Deviation
Flag Outliers	Compute Population Outliers		True
		Minimum Population	15
		IQRatio	1.42

Table 8 Default settings for GE2-nonAT_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		Background IQRatio	1.42
		Use Qtest for Small Populations?	False
	Compute Non Uniform Outliers		True
	Automatically Compute OL Polynomial Terms		False
		Feature – (%CV)^2	0.11000
		Poissonian Noise Term	320
		Background Term	600
		Background – (%CV)^2	0.09000
		Poissonian Noise Term	320
		Background Term	600
Compute Bkgd, Bias and Error	Background Subtraction Method		Minimum Signal (Feature) on Array
	Significance (for IsPosAndSignif and IsWellAboveBG)		Use Pixel Statistics for Significance
		2-sided t-test of feature vs background max p-value	0.01
		WellAboveMulti	2.6
	Signal Correction—Calculate Surface Fit (required for Spatial Detrend)		True
		Feature Set for Surface Fit	AllFeatureTypes
		Perform Filtering for Surface Fit	True
		Perform Spatial Detrending	False
	Signal Correction—Adjust Background Globally		True
		Adjust Background Globally to:	0
	Robust Neg Ctrl Stats?		False
	Choose universal error, or most conservative		Most Conservative
		MultErrorGreen	0.09000

Table 8 Default settings for GE2-nonAT_95 protocol

Protocol Step	Parameter		Default Setting/Value (v.9.5)
		MultErrorRed	0.09000
		Auto Estimate Add Error Red	False
		Additive Error Value Red	30
		Auto Estimate Add Error Green	False
		Additive Error Value Green	30
	Use Surrogates		True
Correct Dye Biases	Dye Normalization Probe Selection Method		Use Rank Consistent Probes
		Rank Tolerance	0.050
		Variable Rank Tolerance	False
		Max Number Ranked Probes	-1
	Omit Background Population Outliers		False
	Allow Positive and Negative Controls		False
	Signal Characteristics		OnlyPositiveAndSignificantSignals
	Normalization Correction Method		Lowess Only
Compute Ratios	Peg Log Ratio Value		4.00
Calculate Metrics	Spikein Target Used		False
	Min Population for Replicate Stats?		5
	PValue for Differential Expression		0.010000
	Percentile Value		75
Generate Results	Generate Single Text File		True
	JPEG Down Sample Factor		4

Differences in protocol settings based on each step

Some of the default settings are the same for all the protocols; yet, many are different, depending on the protocol step.

The table below shows each protocol step and where you can find information on the default settings for that step.

Table 9 Location of protocol template default settings for each step

Protocol Step	Location of default settings
Place Grid	page 46
Find Spots	page 47
Flag Outliers	page 48
Compute Bkgd, Bias and Error	page 51
Correct Dye Biases	page 54
Compute Ratios	page 55
Calculate Metrics	page 55
Generate Results	page 55

Place Grid

The parameters and values for placing the grid are the same for every microarray type, such as GE1, GE2 and CGH. They also appear to be the same for all microarray formats. In fact, they differ depending on the format, but you can't see the differences because the values are hidden.

Formats recognized by the Place Grid algorithm

Recognized Formats
Single Density (11k, 22k)
Double Density (44k)
95k
185k
244k
185k, 10u
244k, 10u
25k
Third Party

When the software automatically determines the format based on the image file or if you select any one of the above formats, the default placement method is Allow Some Distortion. You can also choose Place and Rotate Only. The hidden parameters and values for these two methods differ depending on the format determined or selected.

Find Spots

The parameters and values differ depending on the microarray format.

Table 10 Find Spots – Default values in common and differences for spot formats

Parameter	Default Values	Formats Using Default Value
Use the Nominal Diameter from the Grid Template	True	All
Spot Deviation Limit	1.50	All
Calculation of Spot Statistics Method	Use Cookie	All
Cookie Percentage	0.650	SD, 25k, TP
	0.561	DD, 95k
	0.700	185k, 244k
Exclusion Zone Percentage	1.200	All
Auto Estimate the Local Radius	True	All except 185k, 244k
LocalBGRadius	When False is the default, 100	185k, 244k
	When False is not the default, 127	All except 185k, 244k
Pixel Outlier Rejection Method	Inter Quartile Region	All
RejectIQRFeat	1.42	All
RejectIQRBG	1.42	All
Statistical Method for Spot Values from Pixels	Use Mean/Standard Deviation	All

Flag Outliers

These parameters and values differ depending on the scanner used for the image, the microarray type and the lab protocol.

Table 11 Flag Outliers – Default values in common and differences for protocols

Parameter		Default Values	Protocols Using Default Value
Compute Population Outliers		True	All
	Minimum Population	10	All except GE2-nonAT
		15	GE2-nonAT
	IQRatio	1.42	All
	Background IQRatio	1.42	All
	Use Qtest for Small Populations?	True	GE1-v5, GE2-v5, CGH
		False	GE1-v1, GE2-v4, GE2-SSPE, GE2-nonAT
Compute Non Uniform Outliers		True	All
Agilent scanner			
Automatically Compute OL Polynomial Terms		True	GE1-v1, GE1-v5, GE2-v4, GE2-v5, CGH
		False	GE2-SSPE, GE2-nonAT
	Feature – (%CV)^2	0.01690	GE1-v1, GE2-v4
		0.04000	GE1-v5, GE2-v5, CGH
		.00810	GE2-SSPE
		.11000	GE2-nonAT
(If Automatically Compute is True)	Red Poissonian Noise Term Multiplier	40	GE2-v4
		20	GE2-v5
		5	CGH
	Red Signal Constant Term Multiplier	1	GE2-v4, GE2-v5, CGH

Table 11 Flag Outliers – Default values in common and differences for protocols

Parameter		Default Values	Protocols Using Default Value
	Green Poissonian Noise Term Multiplier	40	GE1-v1, GE2-v4
		20	GE1-V5, GE2-V5
		5	CGH
	Green Signal Constant Term Multiplier	1	GE1-v1, GE1-v5, GE2-v4, GE2-v5, CGH
If Automatically Compute is False	Poissonian Noise Term	320 (R, G combined)	GE2-SSPE, GE2-nonAT
	Background Term	600 (R, G combined)	GE2-SSPE, GE2-nonAT
	Background – (%CV)^2	0.09000	All except GE2-SSPE
		.2250	GE2-SSPE
If Automatically Compute is True	Red Poissonian Noise Term Multiplier	3	GE2-v4, GE2-v5, CGH
	Red Signal Constant Term Multiplier	1	GE2-v4, GE2-v5, CGH
	Green Poissonian Noise Term Multiplier	3	GE1-v1, GE1-v5, GE2-v4, GE2-v5, CGH
	Green Background Constant Term Multiplier	1	GE1-v1, GE1-v5, GE2-v4, GE2-v5, CGH
If Automatically Compute is False	Poissonian Noise Term	320 (R, G combined)	GE2-SSPE, GE2-nonAT
	Background Term	600 (R, G combined)	GE2-SSPE, GE2-nonAT
GenePix (Axon) scanner			
Automatically Compute OL Polynomial Terms		True	All except GE2-nonAT, where not applicable
	Feature – (%CV)^2	0.01440	All
	Red Poissonian Noise Term Multiplier	30	All except GE1 protocols

Table 11 Flag Outliers – Default values in common and differences for protocols

Parameter		Default Values	Protocols Using Default Value
	Red Signal Constant Term Multiplier	1	All except for GE1 protocols
	Green Poissonian Noise Term Multiplier	20	GE1-v1, GE1-v5, GE2-v4, GE2-v5, GE2-SSPE
		15	CGH
	Green Signal Constant Term Multiplier	1	All
	Background – (%CV)^2	0.09000	All
	Red Poissonian Noise Term Multiplier	1	All except GE1 protocols
	Red Background Constant Term Multiplier	1	All except GE1 protocols
	Green Poissonian Noise Term Multiplier	1	All
	Green Background Constant Term Multiplier	1	All

Compute Bkgd, Bias and Error

These parameters and values differ depending on the microarray type and the lab protocol.

Table 12 Compute Bkgd, Bias and Error – Default values in common and differences for protocols

Parameter		Default Values	Protocols Using Default Value
Background Subtraction Method		No Background Subtraction	All except for GE2-nonAT
		Minimum Signal (Feature) on Array	GE2-nonAT
Significance		Use Pixel Statistics for Significance	GE1-v1, GE2-v4, GE2-SSPE, GE2-nonAT
		Use Error Model for Significance	GE1-v5, GE2-v5, CGH
	2-sided t-test of feature vs background max p-value	0.01 (All)	All
	WellAboveMulti	2.6 (All)	GE1-v1, GE2-v4, GE2-SSPE, CGH, GE2-nonAT
		13	GE1-v5, GE2-v5
Signal Correction—Calculate Surface Fit (required for Spatial Detrend)		True	All
	Feature Set for Surface Fit	FeaturesInNegativeControlRange	GE1-v1, GE1-v5, GE2-v4, GE2-v5
		AllFeatureTypes	GE2-SSPE, GE2-nonAT
		Only NegativeControl Features	CGH
	Perform Filtering for Surface Fit	False	GE1-v1, GE2-v4, CGH
		True	GE1-v5, GE2-v5, GE2-SSPE, GE2-nonAT
	Perform Spatial Detrending	True	All except GE2-nonAT
		False	GE2-nonAT
Signal Correction—Adjust Background Globally		False (All except GE2-nonAg)	All except GE2-nonAT

Table 12 Compute Bkgd, Bias and Error – Default values in common and differences for protocols

Parameter		Default Values	Protocols Using Default Value
		True	GE2-nonAT
	Adjust Background Globally to:	0 (When True)	GE2-nonAT
Signal Correction—Perform Multiplicative Detrending		True	GE1-v1, GE1-v5, GE2-v5, CGH
		False	GE2-v4, GE2-SSPE
	Detrend on Replicates Only	False	GE1-v1, CGH
		True	GE1-v5, GE2-v5
	Filter Low signal probes from Fit?	True	GE1-v1, GE1-v5, GE2-v5, CGH
	Neg. Ctrl. Threshold Mult. Detrend Factor	3	GE1-v1, CGH
		5	GE1-v5, GE2-v5
	Perform Filtering for Fit	User Window Average	GE1-v1, GE1-v5, GE2-v5, CGH
	Use polynomial data fit instead of LOESS?	True	GE1-v1, GE1-v5, CGH
	Polynomial Multiplicative DetrendDegree	2	GE1-v1
		4	GE1-v5, CGH
Robust Neg Ctrl Stats?		False	All but CGH
		True	CGH
Choose universal error, or most conservative		Most Conservative	All
	MultErrorGreen	0.1000 ()	GE1-v1,GE1-v5, GE2-v4, GE2-v5, CGH
		0.0800	GE2-SSPE
		0.0900	GE2-nonAT
	MultErrorRed	0.1000 (GE2-v4, CGH-v4)	GE2-v4, GE2-v5, CGH

Table 12 Compute Bkgd, Bias and Error – Default values in common and differences for protocols

Parameter		Default Values	Protocols Using Default Value
		0.0800	GE2-SSPE
		0.0900	GE2-nonAT
	Auto Estimate Add Error Red	True	All except GE1 protocols and GE2-nonAT
		False	GE2-nonAT
	Additive Error Value Red	30 (When False)	GE2-nonAT
	Auto Estimate Add Error Green	True (All except GE2-nonAg)	All except GE2-nonAT
		False	GE2-nonAT
	Additive Error Value Green	30 (When False)	GE2-nonAT
Use Surrogates		True	All except for GE1-v1
		False	GE1-v1

Correct Dye Biases

These parameters and values differ depending on the microarray type. *The 1-color protocol does not correct for dye biases.*

Table 13 Correct Dye Biases – Default values in common and differences for protocols

Parameter	Default Values	Protocols Using Default Values (No GE1 protocols)
Dye Normalization Probe Selection Method	Use Rank Consistent Probes	All
Rank Tolerance	0.050	All
Variable Rank Tolerance	False	All
Max Number Ranked Probes	-1	All except for GE2-v5
	8000	GE2-v5
Omit Background Population Outliers	False	All
Allow Positive and Negative Controls	False (All except for GE1)	All
Signal Characteristics	OnlyPositiveAndSignificantSignals (All except for GE1)	All
Normalization Correction Method	Linear and Lowess	GE2-v4, GE2-v5, GE2-SSPE
	Linear	CGH
	Lowess	GE2-nonAT

Compute Ratios, Calculate Metrics and Generate Results

All of these parameters and values are the same except for the Min Population for Replicate Statistics for CGH-v4. Also, 1-color protocols do not use Compute Ratios.

Table 14 Values in common and differences in protocols

Protocol Step	Parameter	Default Value (v.9.5)
Compute Ratios	Peg Log Ratio Value	4.00
Calculate Metrics	Spikein Target Used?	True
	Min Population for Replicate Statistics	5 (3 for CGH)
	Maximum difference between Grid and Spot Centroid	10
	PValue for Differential Expression	0.010000
Generate Results	Generate Single Text File	True
	JPEG Down Sample Factor	4

2 QC Report Results

QC reports include statistical results to help you evaluate the reproducibility and reliability of your single microarray data. This chapter describes each of five QC reports – 2-color gene expression, 1-color gene expression, CGH, microRNA (miRNA) and 2-color nonAgilent – and how each can help you interpret the performance of your microarray system. Use plots and statistics from the report to:

- Set up your own run charts of statistical values versus time or experiment number to track performance of one microarray compared to other microarrays
- Monitor upstream lab protocols, such as performance of your hybridization/washing steps
- Monitor the effect of changing FE protocol parameters on the performance of your data analysis

If you incorporate a set of QC metrics in your extraction, those results will appear on the final page of the QC report as an Evaluation Table.

QC Reports—Big Picture

2-color Gene Expression QC Report

This module shows you the organization of the 2-color gene expression QC report. See the figure below and on the next pages for links to information on the QC Report regions.

"QC Report Headers" on page 72

1 "Spot Finding of Four Corners" on page 74

2 "Outlier Stats" on page 75

3 "Spatial Distribution of All Outliers" on page 75

4 "Net Signal Statistics" on page 77

5 "Plot of Background-Corrected Signals" on page 79

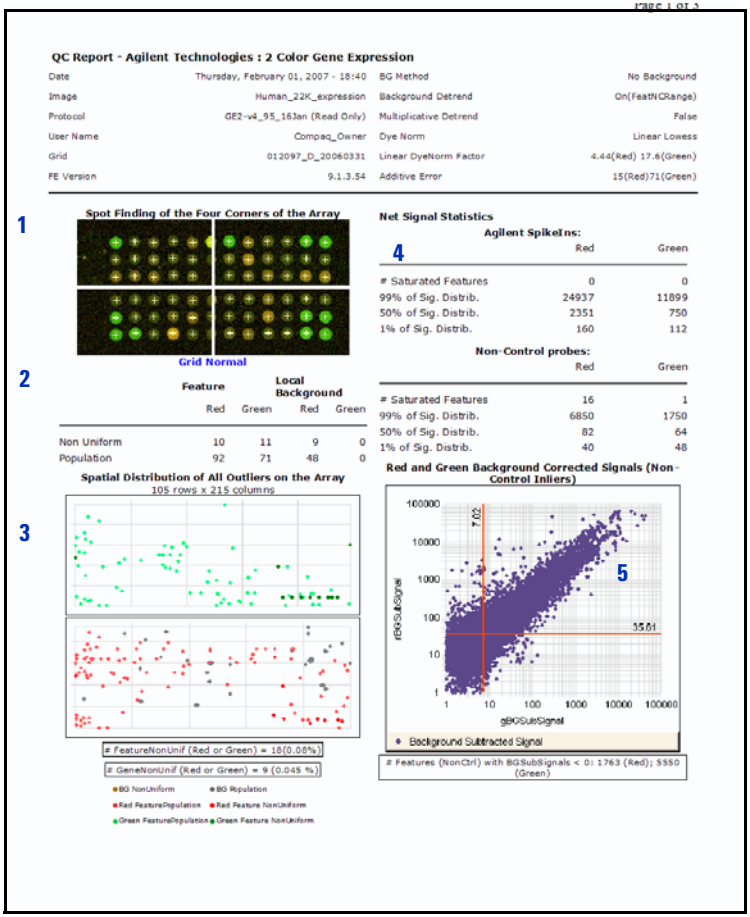
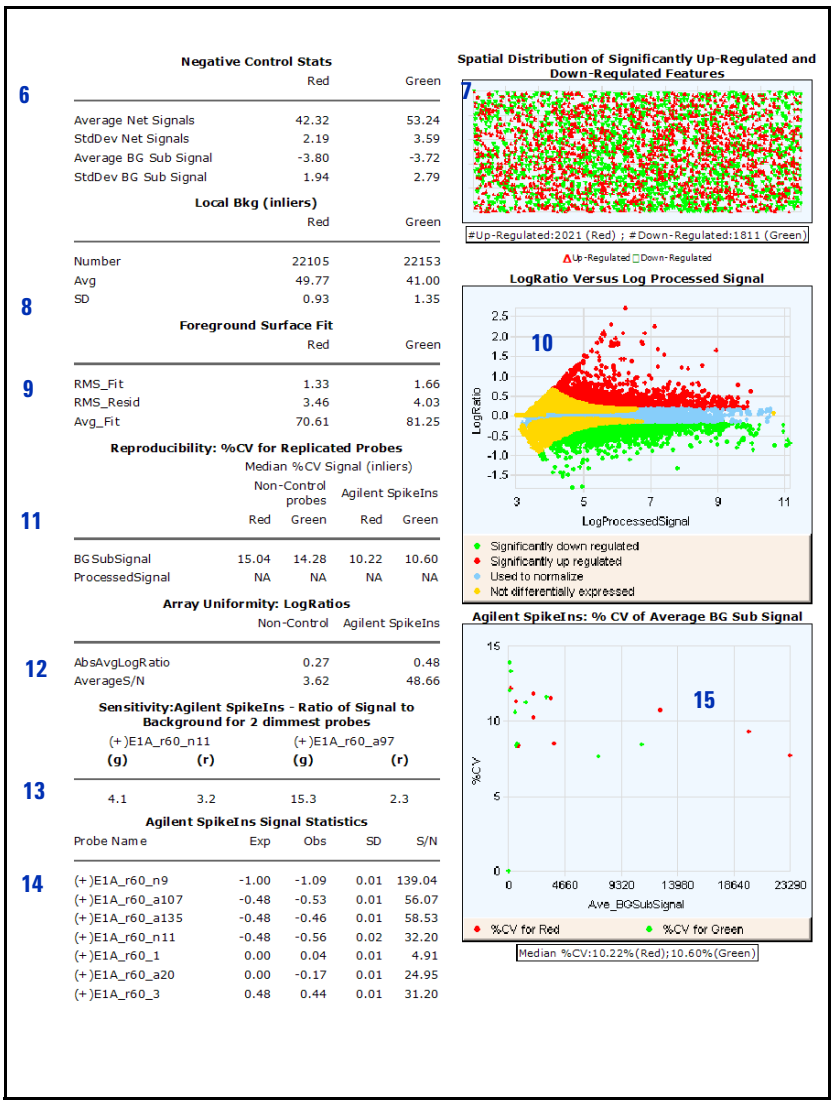


Figure 1 2-color Gene Expression QC Report with Spike-ins (p1)



Negative Control Stats

	Red	Green
Average Net Signals	42.32	53.24
StdDev Net Signals	2.19	3.59
Average BG Sub Signal	-3.80	-3.72
StdDev BG Sub Signal	1.94	2.79

Local Bkg (inliers)

	Red	Green
Number	22105	22153
Avg	49.77	41.00
SD	0.93	1.35

Foreground Surface Fit

	Red	Green
RMS_Fit	1.33	1.66
RMS_Resid	3.46	4.03
Avg_Fit	70.61	81.25

Reproducibility: %CV for Replicated Probes

	Median %CV Signal (inliers)			
	Non-Control probes		Agilent SpikeIns	
	Red	Green	Red	Green
BGSubSignal	15.04	14.28	10.22	10.60
ProcessedSignal	NA	NA	NA	NA

Array Uniformity: LogRatios

	Non-Control	Agilent SpikeIns
AbsAvgLogRatio	0.27	0.48
AverageS/N	3.62	48.66

Sensitivity:Agilent SpikeIns - Ratio of Signal to Background for 2 dimmest probes

(+)E1A_r60_n11		(+)E1A_r60_a97	
(g)	(r)	(g)	(r)
4.1	3.2	15.3	2.3

Agilent SpikeIns Signal Statistics

Probe Name	Exp	Obs	SD	S/N
(+)E1A_r60_n9	-1.00	-1.09	0.01	139.04
(+)E1A_r60_a107	-0.48	-0.53	0.01	56.07
(+)E1A_r60_a135	-0.48	-0.46	0.01	58.53
(+)E1A_r60_n11	-0.48	-0.56	0.02	32.20
(+)E1A_r60_1	0.00	0.04	0.01	4.91
(+)E1A_r60_a20	0.00	-0.17	0.01	24.95
(+)E1A_r60_3	0.48	0.44	0.01	31.20

Figure 2 2-color Gene Expression QC Report with Spike-ins (p2)

16 "Spike-in Linearity Check for 2-color Gene Expression" on page 95

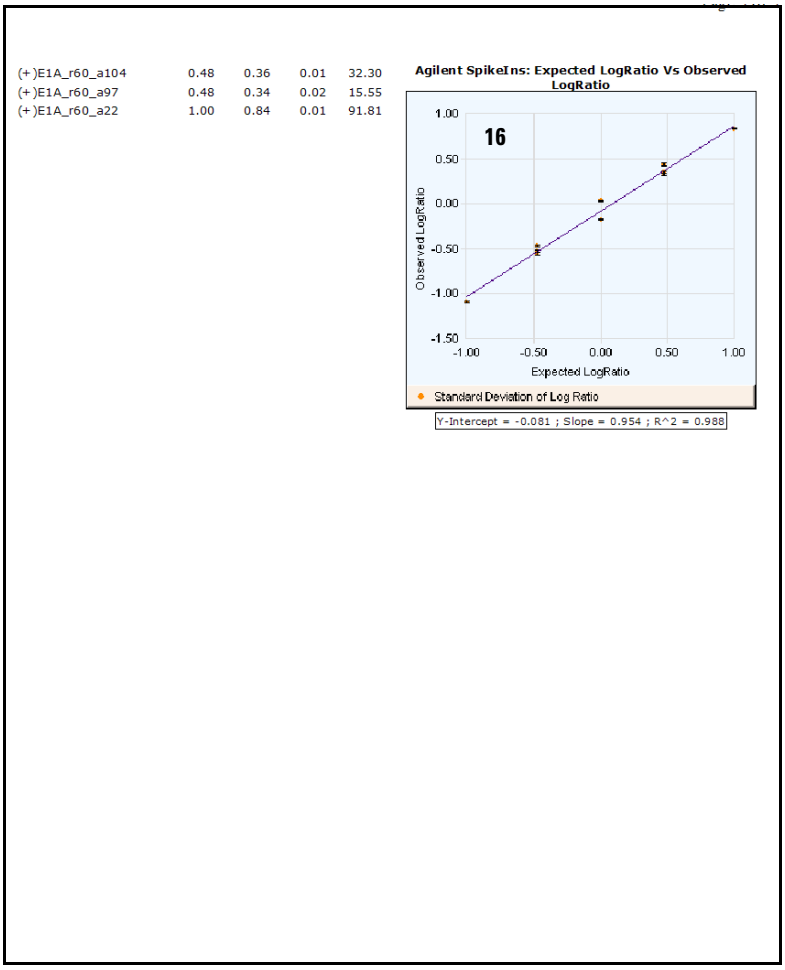


Figure 3 2-color Gene Expression QC Report with Spike-ins (p3)

1-color Gene Expression QC Report

This module shows you the organization of the 1-color gene expression QC report. See the figure below and on the next two pages for links to information on each of the QC Report regions.

"QC Report Headers" on page 72

1 "Spot Finding of Four Corners" on page 74

2 "Outlier Stats" on page 75

3 "Spatial Distribution of All Outliers" on page 75

4 "Net Signal Statistics" on page 77

5 "Histogram of Signals Plot (1-color only)" on page 80

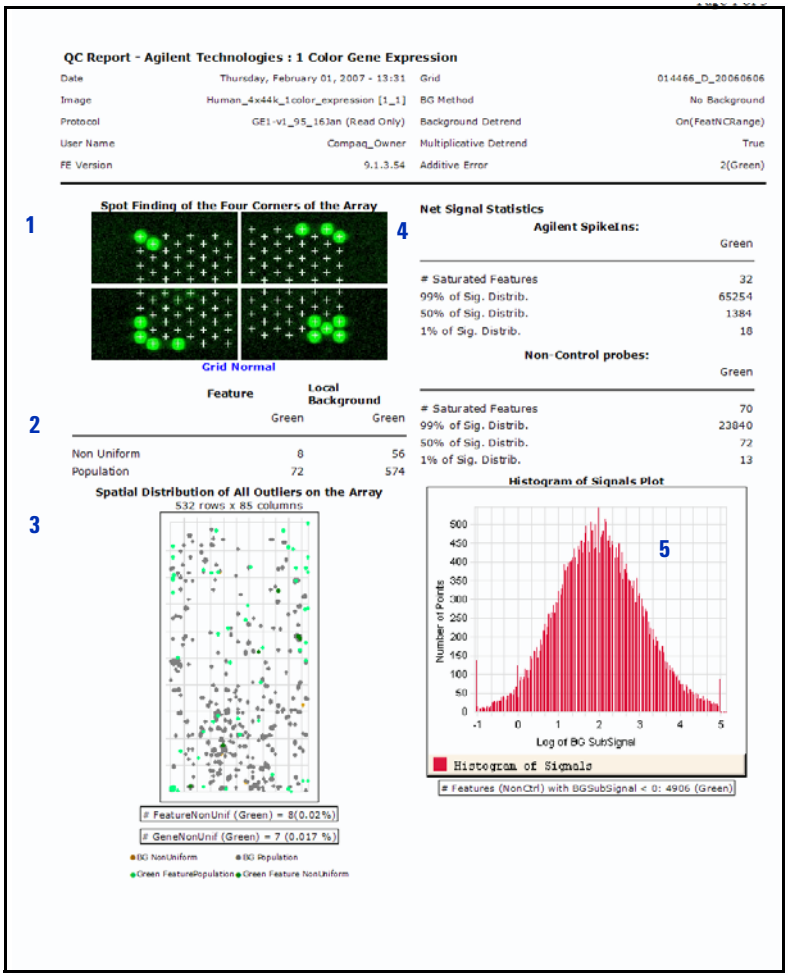


Figure 4 1-color Gene Expression QC Report with Spike-ins (p1)

6 "Negative Control Stats" on page 78

7 "Local Background Inliers" on page 80

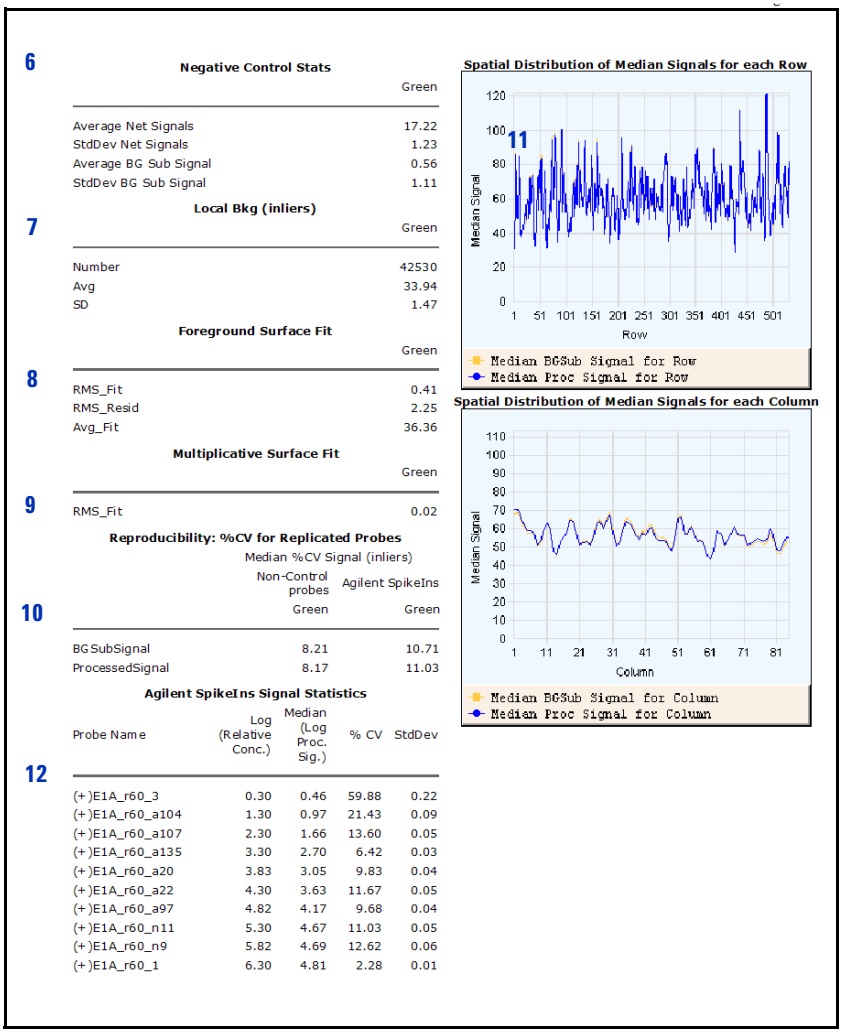
8 "Foreground Surface Fit" on page 81

9 "Multiplicative Surface Fit" on page 82

10 "Reproducibility Statistics (%CV Replicated Probes)" on page 86

11 "Spatial Distribution of Median Signals for each Row and Column (1-color only)" on page 85

12 "1-color gene expression spike-in signal statistics" on page 94



6

Negative Control Stats

	Green
Average Net Signals	17.22
StdDev Net Signals	1.23
Average BG Sub Signal	0.56
StdDev BG Sub Signal	1.11

7

Local Bkg (inliers)

	Green
Number	42530
Avg	33.94
SD	1.47

Foreground Surface Fit

8

	Green
RMS_Fit	0.41
RMS_Resid	2.25
Avg_Fit	36.36

Multiplicative Surface Fit

9

	Green
RMS_Fit	0.02

Reproducibility: %CV for Replicated Probes

10

	Median %CV Signal (inliers)	
	Non-Control probes	Agilent SpikeIns
	Green	Green
BGSubSignal	8.21	10.71
ProcessedSignal	8.17	11.03

Agilent SpikeIns Signal Statistics

12

Probe Name	Log (Relative Conc.)	Median (Log Proc. Sig.)	% CV	StdDev
(+)E1A_r60_3	0.30	0.46	59.88	0.22
(+)E1A_r60_a104	1.30	0.97	21.43	0.09
(+)E1A_r60_a107	2.30	1.66	13.60	0.05
(+)E1A_r60_a135	3.30	2.70	6.42	0.03
(+)E1A_r60_a20	3.83	3.05	9.83	0.04
(+)E1A_r60_a22	4.30	3.63	11.67	0.05
(+)E1A_r60_a97	4.82	4.17	9.68	0.04
(+)E1A_r60_n11	5.30	4.67	11.03	0.05
(+)E1A_r60_n9	5.82	4.69	12.62	0.06
(+)E1A_r60_1	6.30	4.81	2.28	0.01

Figure 5 1-color Gene Expression QC Report with Spike-ins (p2)

13a "Reproducibility plot for 1-color gene expression (spike-in probes)" on page 91

13b "Spike-in Linearity Check for 1-color Gene Expression" on page 96

14 "Table of Values for Concentration-Response Plot (1-color only)" on page 97

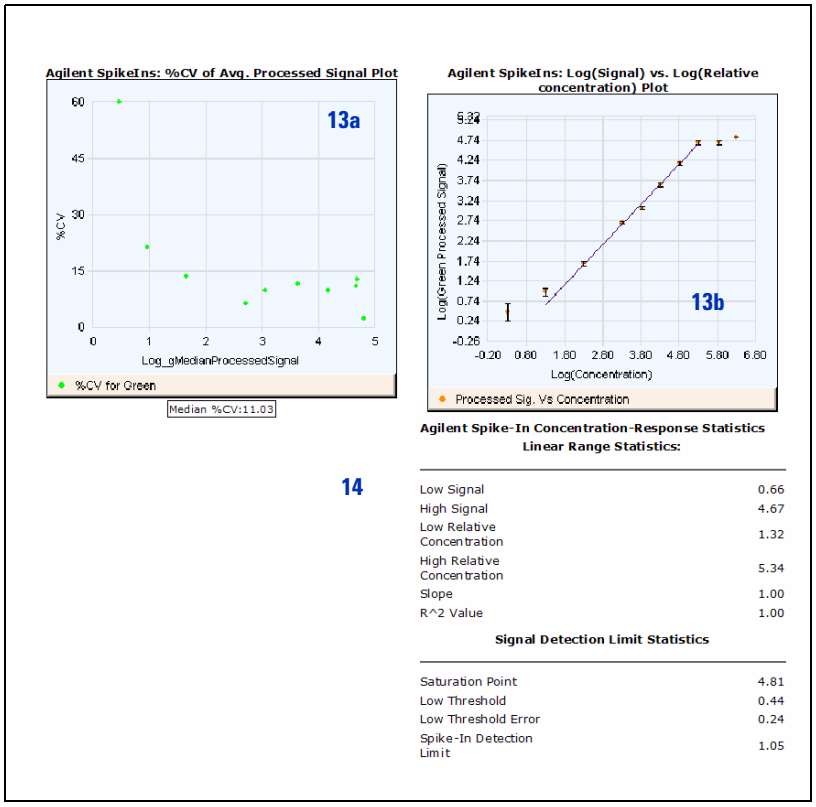


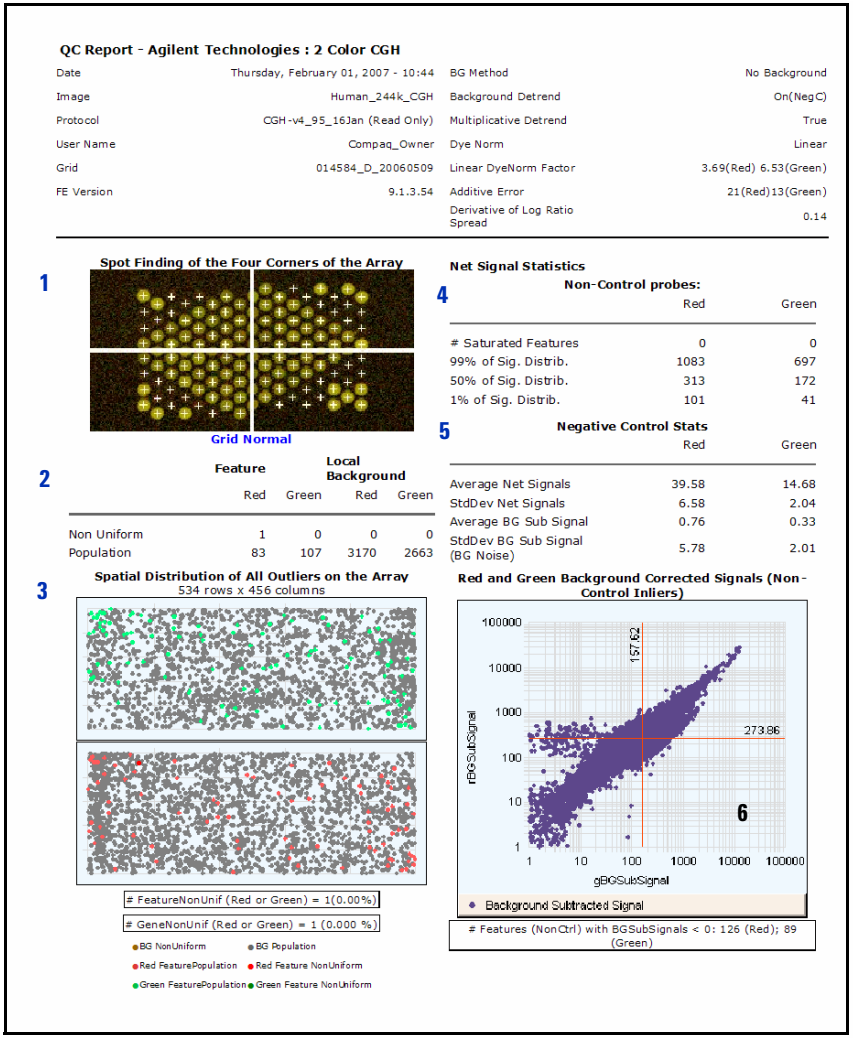
Figure 6 1-color Gene Expression QC Report with Spike-ins (p3)

CGH QC Report

Derivative of Log Ratio Spread is added to the header. See "QC Report Headers" on page 72.

This report lists all of the same information as the 2-color report but removes the Array Uniformity table and spike-ins.

Also, all of the log plots use log base 2 (not 10).



"QC Report Headers" on page 72

1 "Spot Finding of Four Corners" on page 74

2 "Outlier Stats" on page 75

3 "Spatial Distribution of All Outliers" on page 75

4 "Net Signal Statistics" on page 77

5 "Negative Control Stats" on page 78

6 "Plot of Background-Corrected Signals" on page 79

Figure 7 CGH QC Report (Page 1)

7 "Local Background Inliers" on page 80

8 "Foreground Surface Fit" on page 81

9 "Reproducibility Statistics (%CV Replicated Probes)" on page 86

10 "Spatial Distribution of Up-Regulated and Down-Regulated Features (Positive and Negative Log Ratios)" on page 83

11 "Plot of LogRatio vs LogProcessed Signal" on page 84

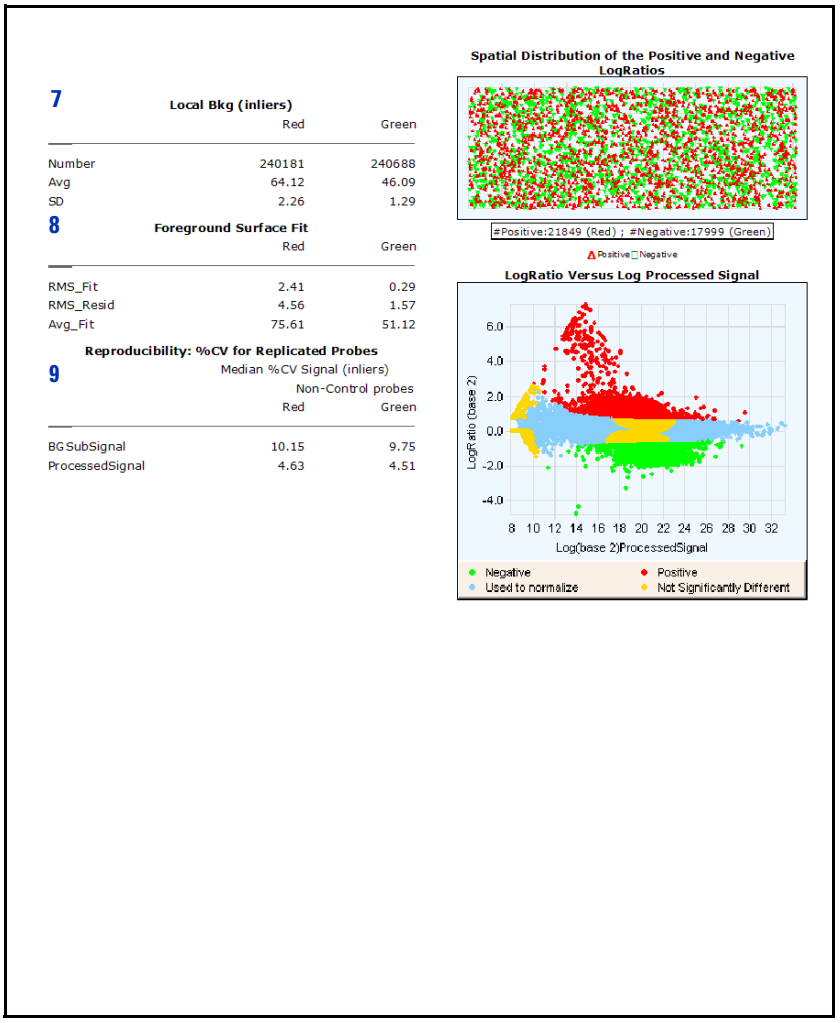


Figure 8 CGH QC Report (Page 2)

MicroRNA (miRNA) QC Report

Agilent miRNA microarrays are currently in development. Please check the Agilent website for the latest information.

This module shows you the organization of the 1-color miRNA QC report. See the figure below and on the next page for links to information on each of the QC Report regions.

"QC Report Headers" on page 72

1 "Spot Finding of Four Corners" on page 74

2 "Outlier Stats" on page 75

3 "Spatial Distribution of All Outliers" on page 75

4 "Net Signal Statistics" on page 77

5 "Negative Control Stats" on page 78

6 "Histogram of Signals Plot (1-color only)" on page 80

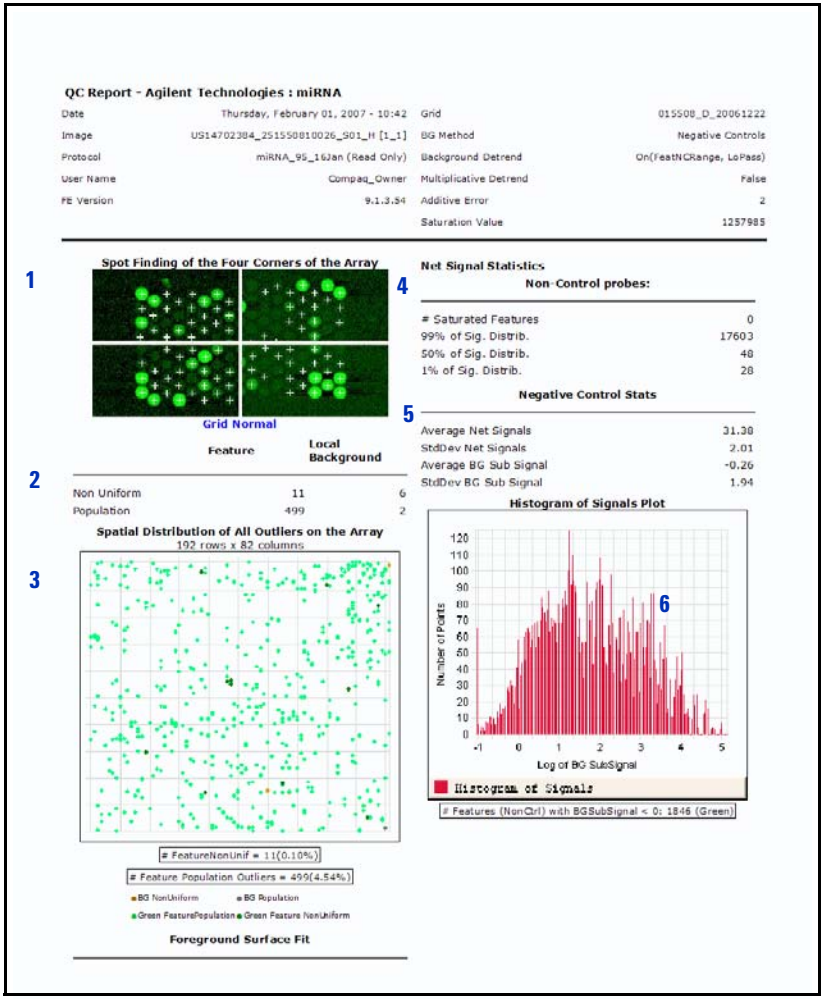


Figure 9 MicroRNA (miRNA) QC Report (p1)

7 "Foreground Surface Fit" on page 81

8 "Reproducibility Statistics (%CV Replicated Probes)" on page 86

9 "Reproducibility plot for miRNA (non-control probes)" on page 92

10 "Spatial Distribution of Median Signals for each Row and Column (1-color only)" on page 85

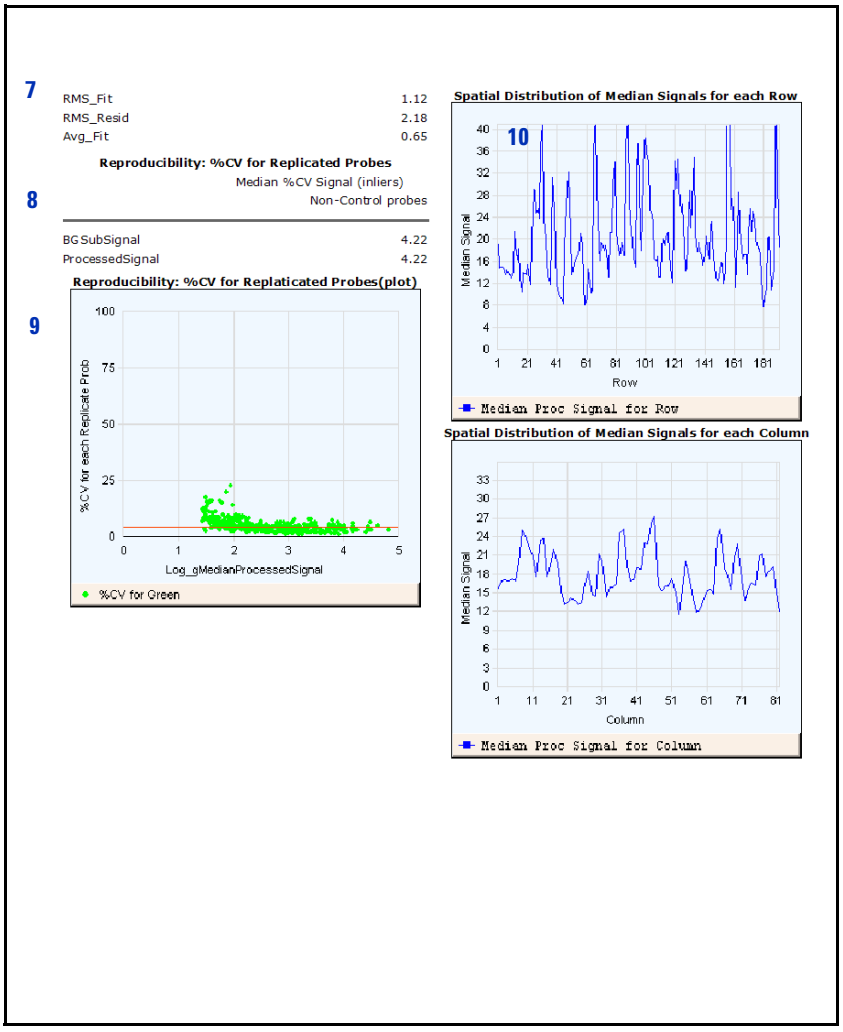
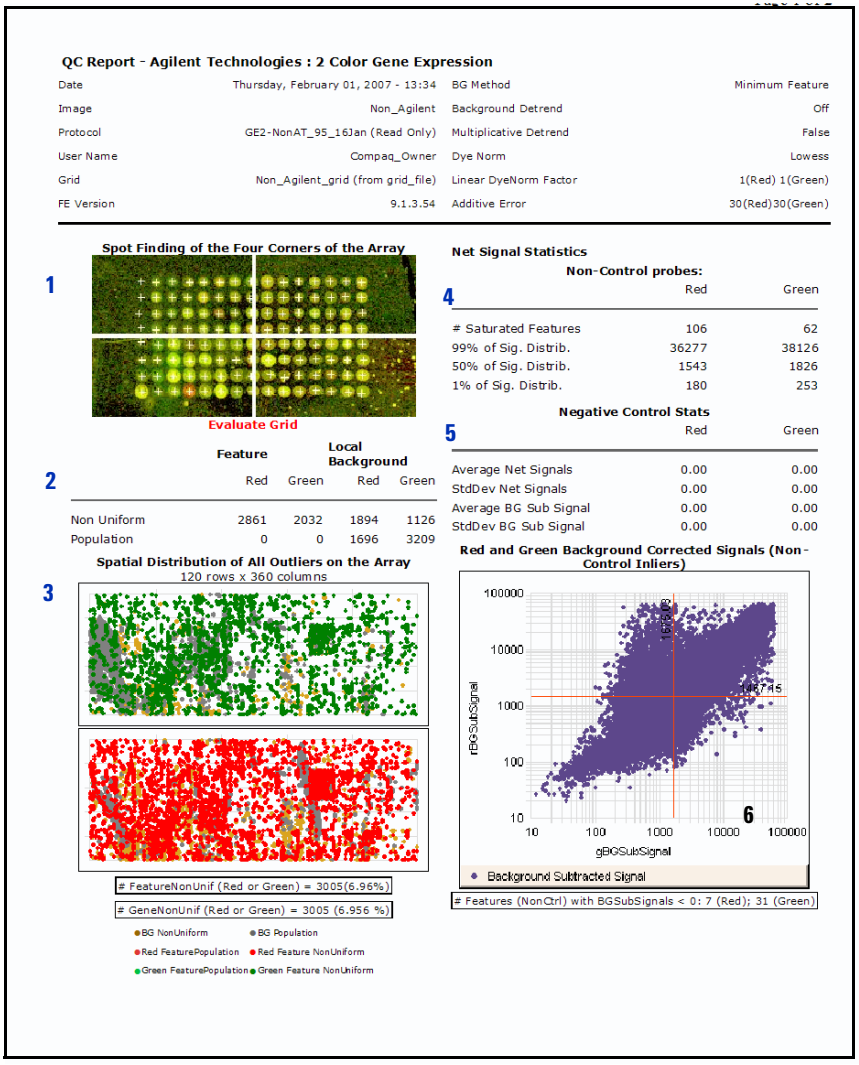


Figure 10 MicroRNA (miRNA) QC Report (p2)

Non-Agilent GE2 QC Report

This report lists all of the same information as the 2-color gene expression QC report but with no spike-ins.



"QC Report Headers" on page 72

1 "Spot Finding of Four Corners" on page 74

2 "Outlier Stats" on page 75

3 "Spatial Distribution of All Outliers" on page 75

4 "Net Signal Statistics" on page 77

5 "Negative Control Stats" on page 78

6 "Plot of Background-Corrected Signals" on page 79

Figure 11 Non-Agilent GE 2 QC Report (Page 1)

7 "Local Background Inliers" on page 80

8 "Foreground Surface Fit" on page 81

9 "Reproducibility Statistics (%CV Replicated Probes)" on page 86

10 "Microarray Uniformity (2-color only)" on page 88

11 "Spatial Distribution of Up-Regulated and Down-Regulated Features (Positive and Negative Log Ratios)" on page 83

12 "Plot of LogRatio vs LogProcessed Signal" on page 84

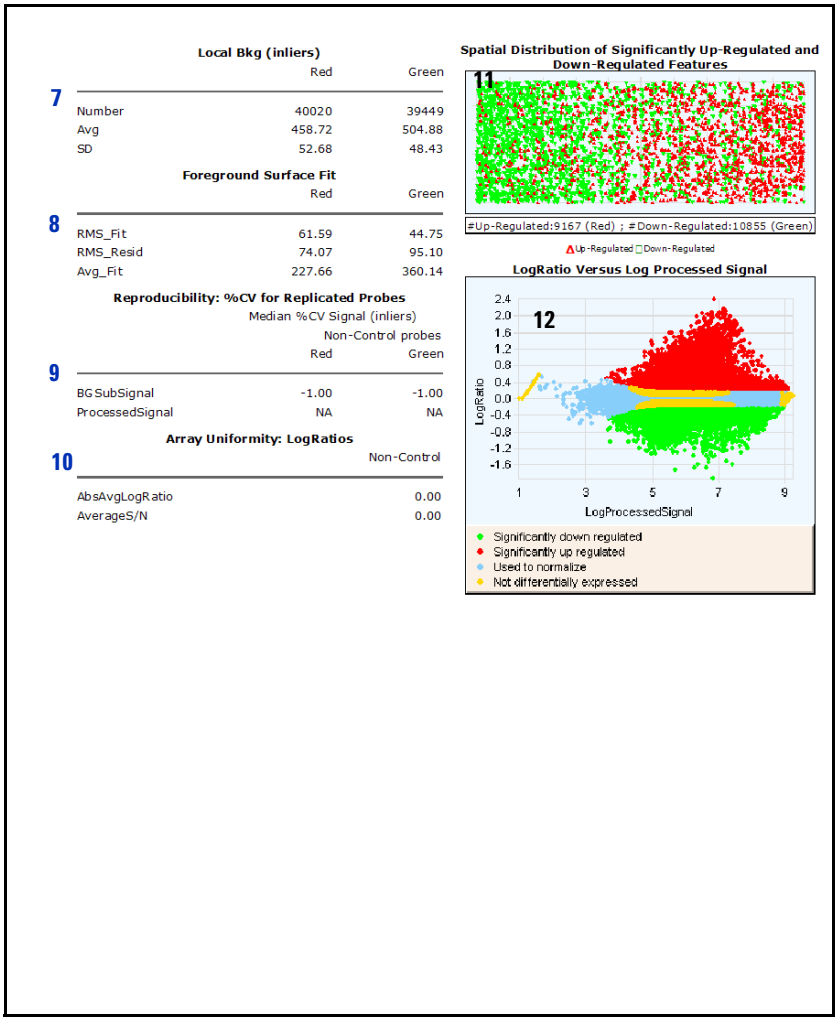


Figure 12 Non-Agilent GE2 QC Report (Page 2)

QC reports with metric sets added

When metric sets are added to the feature extraction, QC reports are generated with an additional set of evaluation metrics. With 2-color gene expression extractions, thresholds are included in the metric set. The evaluation tables then show which thresholds have been exceeded.

Agilent-supplied QC Metric Sets and Thresholds are intended to assist users in monitoring microarray processing issues. They were not developed to detect microarray manufacturing issues.

QC metric set results--default protocol settings

Below is an example of part of a QC report – the header and the Evaluation Table – generated from a 2-color gene expression extraction whose GE2 metric set with thresholds had been added. In this extraction the default protocol settings were used. Note that all values of the metrics are within the default threshold ranges.

QC Report - Agilent Technologies : 2 Color Gene Expression — QCMetrics InRange (12 of 12)

Date	Thursday, February 08, 2007 - 11:15	BG Method	No Background
Image	Human_22K_expression	Background Detrend	On(FeatNCRange, LoPass)
Protocol	GE2-v5_95_Feb07 (Read Only)	Multiplicative Detrend	True
User Name	Compaq_Owner	Dye Norm	Linear Lowess
Grid	Human_22K_expression_GE2-v5_95_Feb07_grid (from grid_file)	Linear DyeNorm Factor	4.15(Red) 16(Green)
FE Version	9.1.4.57	Additive Error	14(Red)63(Green)

Evaluation Metrics for GE2_QCMT_Feb07

Metric Name	Value	UpLim	LowLim	IsMandatory
AnyColorPrcntBGNonUnifOL	0.04	5.00	NA	False
AnyColorPrcntFeatNonUnif...	0.08	1.00	NA	False
absE1aObsVsExpCorr	0.99	NA	0.86	False
absE1aObsVsExpSlope	0.95	NA	0.85	False
gE1aMedCVBkSubSignal	10.59	25.00	NA	False
gNegCtrlAveBGSubSig	-3.79	10.00	-20.00	False
gNegCtrlSDevBGSubSig	2.83	15.00	NA	False
gNonCntrlMedCVBkSubSignal	14.57	25.00	NA	False
rE1aMedCVBkSubSignal	10.21	25.00	NA	False
rNegCtrlAveBGSubSig	-3.93	4.00	-20.00	False
rNegCtrlSDevBGSubSig	2.07	6.00	NA	False
rNonCntrlMedCVBkSubSignal	15.05	25.00	NA	False

Figure 13 Partial QC Report—Header and Evaluation Metrics with GE2 metric set with thresholds added—Default protocol settings

QC metric set results—Spatial and Multiplicative Detrending Off

Below is an example of a QC report header and Evaluation Table generated from a 2-color gene expression extraction whose GE2 metric set with thresholds had been added. In this extraction spatial and multiplicative detrending were turned off. Note that not all values of the metrics are within the default thresholds.

QC Report - Agilent Technologies : 2 Color Gene Expression — QCMetrics InRange (10 of 12)

Date	Thursday, February 08, 2007 - 11:15	BG Method	No Background
Image	Human_22K_expression	Background Detrend	NA
Protocol	GE2-v5_95_Feb07_AllDTOff (Editable)	Multiplicative Detrend	False
User Name	Compaq_Owner	Dye Norm	Linear Lowess
Grid	Human_22K_expression_GE2-v5_95_Feb07_grid (from grid_file)	Linear DyeNorm Factor	4.05(Red) 6.85(Green)
FE Version	9.1.4.57	Additive Error	4(Red)7(Green)

Evaluation Metrics for GE2_QCMT_Feb07

Metric Name	Value	UpLim	LowLim	IsMandatory
AnyColorPrcntBGNonUnifOL	0.04	5.00	NA	False
AnyColorPrcntFeatNonUnif...	0.08	1.00	NA	False
absE1aObsVsExpCorr	0.97	NA	0.86	False
absE1aObsVsExpSlope	0.87	NA	0.85	False
gE1aMedCVBkSubSignal	7.64	25.00	NA	False
gNegCtrlAveBGSubSig	77.46	10.00	-20.00	False
gNegCtrlSDevBGSubSig	3.59	15.00	NA	False
gNonCntrlMedCVBkSubSignal	4.78	25.00	NA	False
rE1aMedCVBkSubSignal	9.27	25.00	NA	False
rNegCtrlAveBGSubSig	66.53	4.00	-20.00	False
rNegCtrlSDevBGSubSig	2.19	6.00	NA	False
rNonCntrlMedCVBkSubSignal	6.56	25.00	NA	False

♦ In Normal Range ♦ Evaluate

Figure 14 QC Report Header and Evaluation Metrics with GE2 metric set with thresholds added—Detrending turned off

QC Report Headers

2-color Gene Expression QC Report

The following Feature Extraction information is found in the 2-color gene expression QC Report header:

Date	Date and time that the QC Report was generated
Image	Name of the TIFF file that was extracted
Protocol	Name of the protocol used for the extraction
User Name	Name of the user who set up the extraction
Grid	Name of the grid template or grid file used
FE Version	Version of the Feature Extraction software used
BG Method	Type of background subtraction method used
Background Detrend	If Spatial Detrend was turned on or off during the extraction
Multiplicative Detrend	If Multiplicative Detrend was turned on or off during the extraction
Dye Norm	Type of dye normalization method used
Linear DyeNorm Factor	Global dye normalization factor determined for the linear portion of the correction method.
Additive Error	Additive portion of the error estimated in the Universal or Most Conservative error model if AutoEstimateAddError was selected, or the values entered into the protocol, if AutoestimateAddError was not selected.
Saturation Value	The signal intensity value above which the signal is considered saturated. This value only appears if it exceeds about 65,500. If it appears, this means that this QC report is from an XDR image file.

1-color Gene Expression QC Report

This report lists all of the same information as the 2-color gene expression report except for Global Adjust, Dye Norm and Linear DyeNorm Factor, which are removed.

CGH QC Report

All header information that appears in the 2-color gene expression QC report also appears in the CGH report. This report also adds the metric, Derivative of Log Ratio Spread, in the Header information.

Derivative of Log Ratio Spread Measures the standard deviation of the probe-to-probe difference of the log ratios. This is a metric used in CGH experiments where differences in the log ratios are small on average. A smaller standard deviation here indicates less noise in the biological signals.

MicroRNA (miRNA) QC Report

This header lists the same information as the 1-color gene expression QC Report header. It also lists Saturation Values exceeding 65,500 if the XDR function is turned on. Because the dynamic range of the intensity for all miRNA microarray spots on a microarray may exceed that of a normal scan range, the miRNA analysis on some microarrays may benefit with the XDR function turned on.

Non-Agilent 2-color gene expression QC Report

This header lists the same information as the 2-color gene expression QC report header.

Feature Statistics

This section provides an explanation for each of the segments of the QC report that cover feature statistics and how these feature statistics can help you assess the performance of your microarray system.

Spot Finding of Four Corners

By viewing the features in the four corners of the microarray, you can note if the spot centroids have been located properly. If their locations are off-center in one or more corners, you may have to run the extraction again with a new grid.

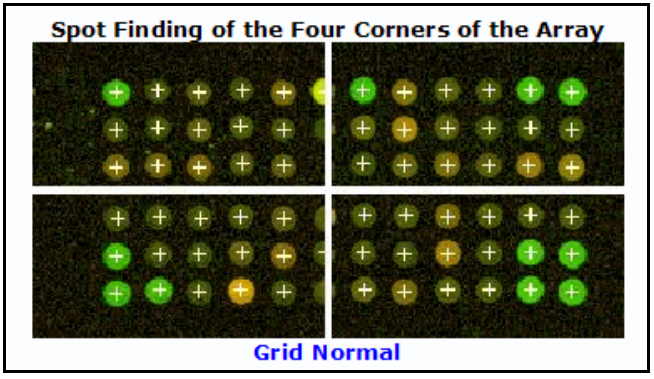


Figure 15 QC Report—Spot Finding for Four Corners

Outlier Stats

If the QC Report shows a greater than expected number of non-uniform or population outliers, you may want to check your hybridization/wash step. Also, check the visual results (.shp file) to see if the spot centroids are off-center. If the grid was not placed correctly, a new grid is required.

	Feature		Local Background	
	Red	Green	Red	Green
Non Uniform	1	5	7	4
Population	50	50	2308	2099

Figure 16 QC Report—Outlier Stats

For 1-color reports, the number of outliers is reported for the green channel only.

Spatial Distribution of All Outliers

The QC report shows two plots of all the outliers, both population and nonuniformity outliers, whose positions are distributed across the microarray. One plot is for the green channel, and the other, for the red channel.

To distinguish the background population and nonuniform outliers from one another, view the color coding at the bottom of the two plots.

For the 1-color report, only the green plot is shown.

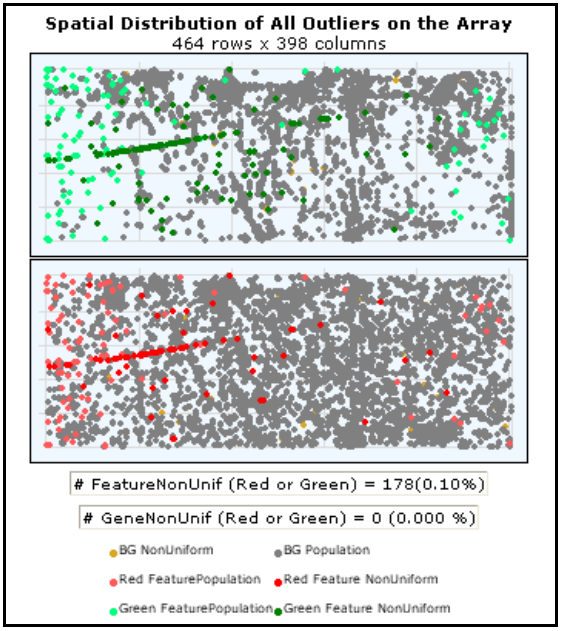


Figure 17 QC Report—Number and Spatial Distribution of Outliers

The number (and percentage) of features that are feature nonuniformity outliers in either the green or red channel is shown below the plot. The 1-color report shows only the percentage of green feature non-uniformity outliers.

Also, the number (and percentage) of genes that are nonuniformity outliers in either channel is shown below the plot. If there were replicate features representing one gene and at least one feature was not an outlier, no gene outliers would appear.

Net Signal Statistics

Net signal is the mean signal minus the scanner offset. Net signal is used so that these statistics are independent of the scanner version.

Net signal statistics are an indication of the dynamic range of the signal on a microarray for both non-control probes and spike-in probes (only non-control probes for the CGH QC report). The QC Report uses the range from the 1st percentile to the 99th percentile as an indicator of dynamic range for that microarray. NetSignal is also a column in the FeatureData output.

For example, in the figure below for non-control probes the dynamic range of the net signal intensity for the red channel is from 42 to 6803 with half the probes having a net signal intensity of greater than the median of 97 and half below the median of 97. The median (or 50th percentile) represents the middle of the ranked-values of the distribution of signals.

Another indicator of signal range for the microarray is the number of features that are saturated in the scanned image (i.e., NumSat).

Net Signal Statistics

Agilent SpikeIns:

	Red	Green
NumSat	0	0
99%	24913	11899
50%	2351	750
1%	160	112

Non-Control probes:

	Red	Green
NumSat	16	1
99%	6850	1750
50%	82	64
1%	40	48

Figure 18 QC Report—Net Signal Statistics

Negative Control Stats

The Negative Control Stats table includes the average and standard deviation of the net signals (mean signal minus scanner offset) and the background-subtracted signals for both the red and green channels in the negative controls. These statistics filter out saturated and feature non-uniform and population outliers and give a rough estimate of the background noise on the microarray.

Negative Control Stats

	Red	Green
Average Net Signals	52.68	51.49
StdDev Net Signals	9.55	9.48
Average BG Sub Signal	2.22	1.95
StdDev BG Sub Signal (BG Noise)	9.25	8.52

Figure 19 QC Report—Negative Control Stats

Plot of Background-Corrected Signals

Figure 20 is a plot of the log of the red background-corrected signal versus the log of the green background-corrected signal for non-control inlier features. The linearity or curvature of this plot can indicate the appropriateness of background method choices. The plot should be linear.

The intersection of the red vertical and horizontal lines shows the location of the median signal. The numbers along the edge of the lines represent the location of the median signal on the plot.

The values below the plot indicate the number of non-control features that have a background-corrected signal less than zero.

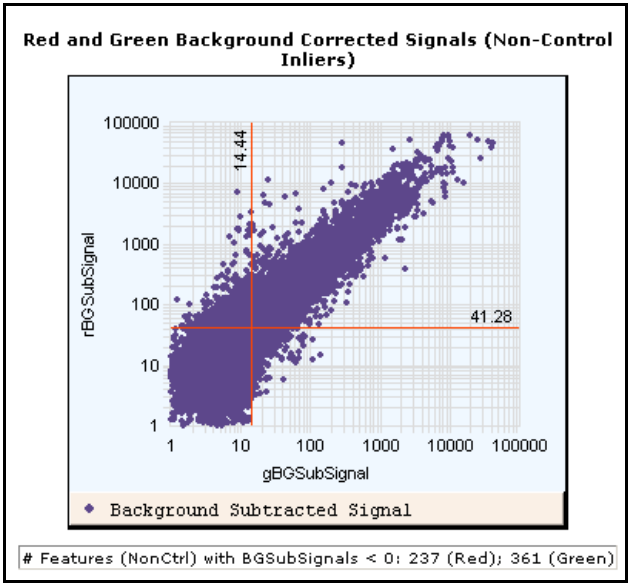


Figure 20 QC Report—Plot of Background-Corrected Signals

Histogram of Signals Plot (1-color only)

The purpose of this histogram is to show the level of signal and the shape of the signal distribution. The histogram is a line plot of the number of points in the intensity bins vs. the log of the processed signal.

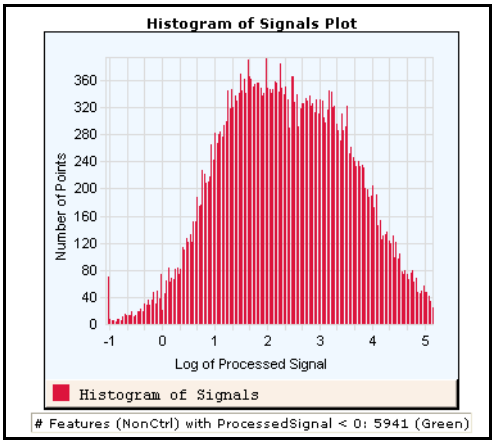


Figure 21 1-color QC Report—Histogram of Signals Plot

Local Background Inliers

With these numbers you can see the mean signal distribution for the local background regions (BGMeanSignal) after outliers have been removed. This information can help you detect hybridization/wash artifacts and can be a component of noise in the low signal range.

Local Bkg (inliers)		
	Red	Green
Number	22104	22153
Avg	49.77	41.00
SD	0.93	1.35

Figure 22 QC Report—Local Background Inliers

Foreground Surface Fit

See "Step 12: Perform background spatial detrending to fit a surface" on page 215 of this guide for more information about these calculations.

Spatial Detrend attempts to account for low signal background that is present on the feature "foreground" and varies across the microarray.

- A high RMS_Fit number can indicate gradients in the low signal range before detrending.
- RMS_Resid indicates residual noise after detrending.
- AvgFit indicates how much signal is in the "foreground".

  A higher AvgFit number indicates a larger amount of signal was detected by the detrend algorithm and removed.

  This value may include the scanner offset, unless a background method has been used before detrending. The value may not include higher frequency background signals. These higher frequency background signals are best removed by using the Local Background Method before the detrending algorithm.

Foreground Surface Fit		
	Red	Green
RMS_Fit	1.54	1.51
RMS_Resid	1.83	1.43
Avg_Fit	64.98	74.00

Figure 23 QC Report—Foreground Surface Fit

Multiplicative Surface Fit

See "Step 15: Determine the error in the signal calculation" on page 221 of this guide for more information about these calculations.

This is the root mean square (RMS) of the surface fit for the data. The RMS X 100 is roughly the average % deviation from "flat" on the microarray. A multiplicative trend means that there are regions of the microarray that are brighter or dimmer than other regions. This trend is an effect that multiplies signals; that is, a brighter signal is more affected in absolute signal counts than a dimmer signal.

This option is turned on in GE1, GE2 (v5) and CGH protocols, turned off in the miRNA protocol and is not available for non-Agilent protocols.

If the signal is not improved through a multiplicative surface fit, then the software turns the algorithm off, and the RMS_Fit shows up as 0.0, as in the figure below.

Multiplicative Surface Fit

	Green
RMS_Fit	0.00

Figure 24 QC Report—Multiplicative Surface Fit

What if multiplicative detrending does not work?

If the median %CV for the Processed Signal of the non-control probes is greater than the BGSub Signal median %CV after multiplicative detrending, Feature Extraction turns off multiplicative detrending.

The QC report shows an RMS_Fit = 0.0 if multiplicative detrending did not result in better data.

If there are no stats for non-control probes, FE looks at the spike-in control probes. If the %CVs for these become worse, FE removes detrending.

If the option "Detrend on Replicates only" is chosen and if there are not enough replicates for non-control or spike-in control probes, FE turns off multiplicative detrending.

Spatial Distribution of Up-Regulated and Down-Regulated Features (Positive and Negative Log Ratios)

You can view the distribution of the significantly up- and down-regulated features on this plot (up–red; down–green).

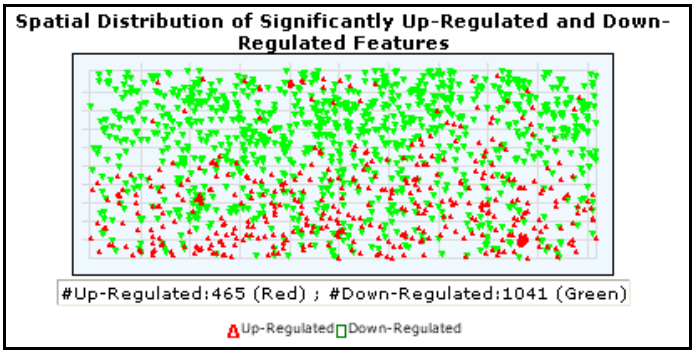


Figure 25 QC Report—Spatial Distribution of Up- and Down-Regulated Features

For the CGH QC Report, these are referred to as “Positive” and “Negative” log ratios.

If the microarray contains greater than 5000 features, the software randomly selects 5000 data points. These points include the number of up-regulated features in the same proportion to the number of down-regulated features as they are found on the actual microarray.

The threshold that is used to determine significance is set in the protocol—QCMetrics_differentialExpressionPValue.

These are the same features shown as up- or down-regulated in Figure 26.

Plot of LogRatio vs LogProcessed Signal

This plot shows the log ratios of non-control inliers vs. the log of their red and green processed signals. The color coding signifies the degree to which features are significantly differentially expressed: those that are up-regulated (red), those that are down-regulated (green) and those that cannot confidently be said to show gene expression (light yellow). For the CGH QC Report, these are referred to as “Positive”, “Negative” log ratios (base 2). The threshold that is used to determine significance is set in the protocol (QCMetrics_differentialExpressionPValue).

Features that were used for normalization are indicated in blue. Significance takes precedence over normalization for the color coding; that is, features that are both significantly differentially expressed and used for normalization will be color-coded either red or green.

LogProcessedSignal in the plot is [Log(rProcessedSignal x gProcessedSignal)]/2.

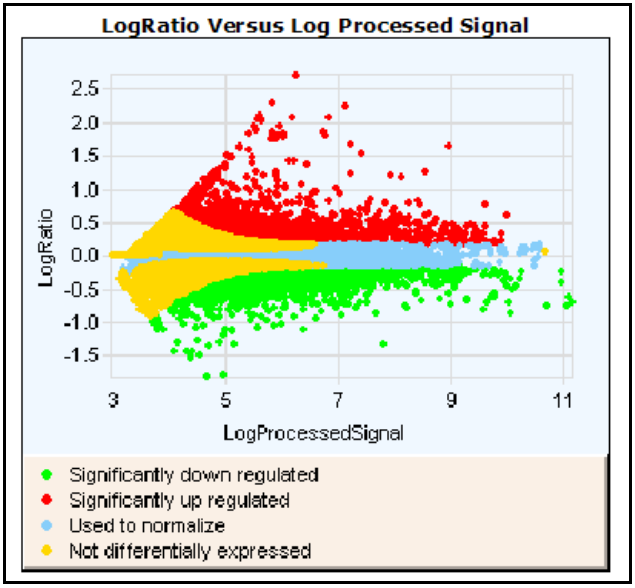


Figure 26 QC Report—Plot of Up- and Down-Regulated Features

Spatial Distribution of Median Signals for each Row and Column (1-color only)

The first of these graphs plots the median Processed Signal and median BGSub Signal for each row over all columns of a 1-color GE microarray. The second plots the same signals for each column over all rows of the 1-color GE microarray. The difference between the Processed Signal and the BGSubSignal represents the effect of the multiplicative detrending. The Processed Signal should look flatter.

Higher frequency noise is shown in these plots so you can distinguish a low frequency trend outside of the high frequency noise.

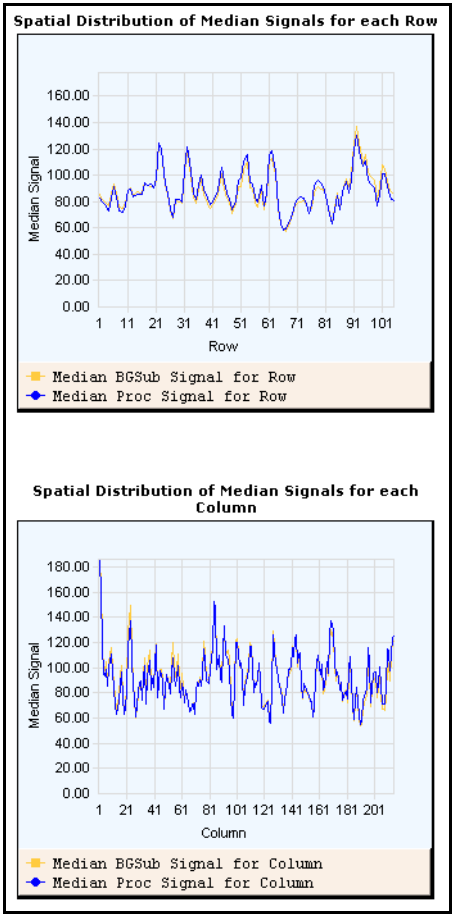


Figure 27 1-color QC Report—Median Signal Spatial Distribution

Inter-Feature Statistics

Spike-in probes are known probes that are hybridized with known quantities of a target "spike-in" cocktail. They are used to perform a quality check of the microarray/experiment.

Some microarray designs have replicated non-control probes; that is, multiple features on the microarray contain the same probe sequence. Many of the Agilent microarray designs also have *spike-in probes*, which are replicated across the microarray (e.g., some microarrays have 10 sequences with 30 replicates each). The QC Report uses these replicated probes to evaluate reproducibility of both the signals and the log ratios. Metrics such as signal %CV and log ratio statistics are calculated if probes are present with a minimum number of replicates.

The protocol indicates if labeled target to these spike-in probes has been added in the hybridization (QCMetrics_UseSpikeIns). The minimum number of replicates (inliers to Sat & NonUnif flagging) is also set in the protocol (QCMetrics_minReplicate Population).

This section provides an explanation for each of the segments of the QC report that cover inter-feature statistics and how these replicate statistics can help you assess performance.

Reproducibility Statistics (%CV Replicated Probes)

Non-control probes

If a non-control probe has a minimum number of inliers, a %CV (percent coefficient of variation) of the background-corrected signal is calculated for each channel (SD of signals/average of signals). This calculation is done for each replicated probe, and the median of those %CV's is reported in the table for each channel.

Reproducibility: %CV for Replicated Probes

	Median %CV Signal (inliers)			
	Non-Control probes		Agilent SpikeIns	
	Red	Green	Red	Green
BGSubSignal	22.36	21.68	21.75	21.16
ProcessedSignal	NA	NA	NA	NA

Figure 28 QC Report—Reproducibility

A lower median %CV value indicates better reproducibility of signal across the microarray than a higher value.

Exclusion of dim probes

Feature Extraction calculates the Median %CV using those probes bright enough to be in the range where the noise is more proportional to signal. FE excludes from the calculation any sequences for which the Average (BGSubSignal) x Multiplicative error < Additive error/Dye Norm Factor. For 1-color data the Dye Norm Factor is 1.

A probe sequence will have a %CV calculated if the number of features that pass the filters (NonUniform and signal filter, described above) is greater than the minimum replicate number indicated in the protocol: “QCMetrics_minReplicatePopulation”.

If the number of replicated sequences with enough inlier features is less than 10 or less than 10% of the replicated sequence, that is, if there are not enough bright replicated probes, the Median %CV field shows up as -1.

Spike-in probes

The same algorithm is used to calculate the Median %CV for the spike-in probes as well. Because there are only 10 sequences in total and some are expected to fail the Additive error test described above, the minimum number of “bright enough” sequences required to calculate the Median %CV is 3.

Microarray Uniformity (2-color only)

The QC Report has two metrics that measure the uniformity of replicated log ratios and that indicate the span of log ratios: average S/N and AbsAvgLogRatio. These are calculated from inlier features of replicated non-control and spike-in probes.

For example, some microarrays have 100 different non-control probe sequences with 10 replicate features each. For each replicate probe, the average and SD of the log ratios are calculated. The signal to noise (S/N) of the log ratio for each probe is calculated as the absolute of the average of the log ratios divided by the SD of the log ratios. From the population of 100 S/N's, for example, the average S/N is determined and shown in the table below.

The second metric, AbsAvgLogRatio, indicates the amount of differential expression (up-regulated or down-regulated). As described above, averages of log ratios are calculated for each replicated probe. The absolute of these averages is determined next. Then, the average of these absolute of averages is calculated to get a single value for the QC Report. The larger this value, the more differential expression is present.

Array Uniformity: LogRatios

	Non-Control	Agilent SpikeIns
AbsAvgLogRatio	0.27	0.47
Average S/N	2.83	52.43

Figure 29 QC Report—Array Uniformity: LogRatios

Sensitivity

These values represent the NetSignal to background (BGUsed - ScannerOffset) ratio of the two spike-in probes with the lowest background-subtracted signal. Their purpose is to characterize the sensitivity of detecting a low signal relative to the background.

Sensitivity:Agilent SpikeIns - Ratio of Signal to Background for 2 dimmest probes

(+)E1A_r60_n11		(+)E1A_r60_a97	
(g)	(r)	(g)	(r)
3.7	4.7	2.6	17.3

Figure 30 QC Report—Sensitivity: Agilent SpikeIns Ratio of Signal to Background for 2 dimmest probes

Reproducibility Plots

Reproducibility plot for 2-color gene expression (spike-in probes)

Signal replicate statistics are calculated for spike-in probes if three criteria are met:

- They are present on the microarray.
- The protocol indicates that labeled target to these spike-in probes has been added in the hybridization (QCMetrics_UseSpikeIns is True).
- There are a minimum number of inlier features for calculations (QCMetrics_minReplicatePopulation).

As described above for non-control probes, %CV's are calculated for inliers for both red and green background-corrected signals. The %CV for each probe is plotted on the next page vs. the average of its background-corrected signal. The median of these %CV's is shown directly beneath the plot.

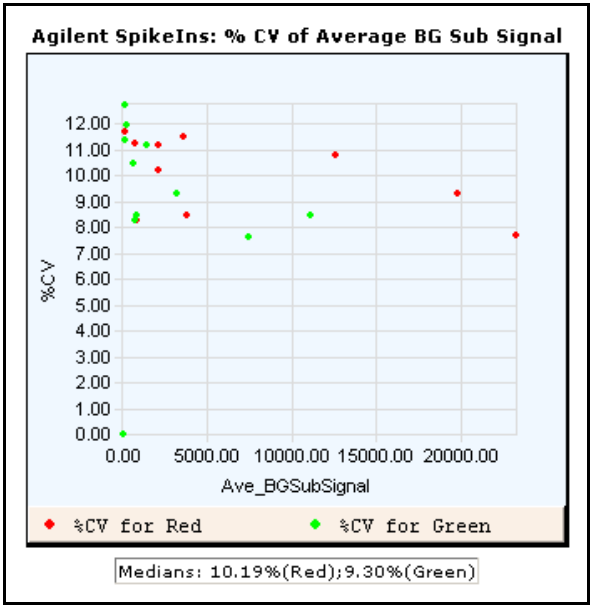


Figure 31 QC Report—Agilent SpikeIns: %CV of Average BGSub Signal

Reproducibility plot for 1-color gene expression (spike-in probes)

This graph plots %CV vs. the log_gMedianProcessedSignal for the 1-color gene expression microarray experiment. The region where the %CV flattens out and is not tightly correlated with signal is the range where noise is proportional to signal. This is generally the range used to calculate the median %CV.

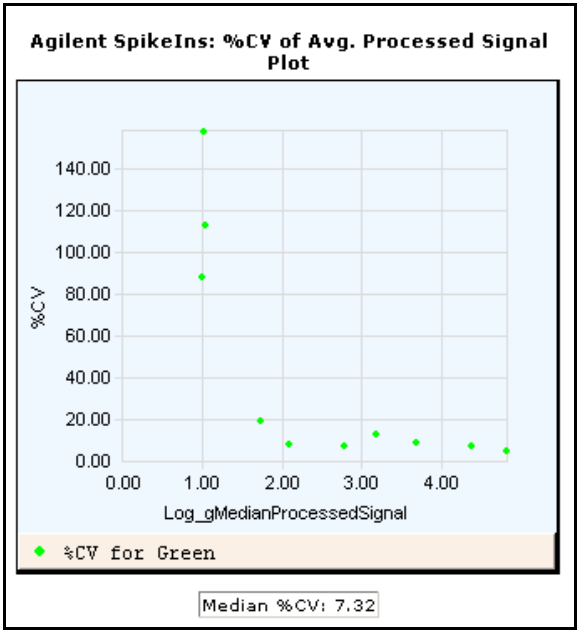
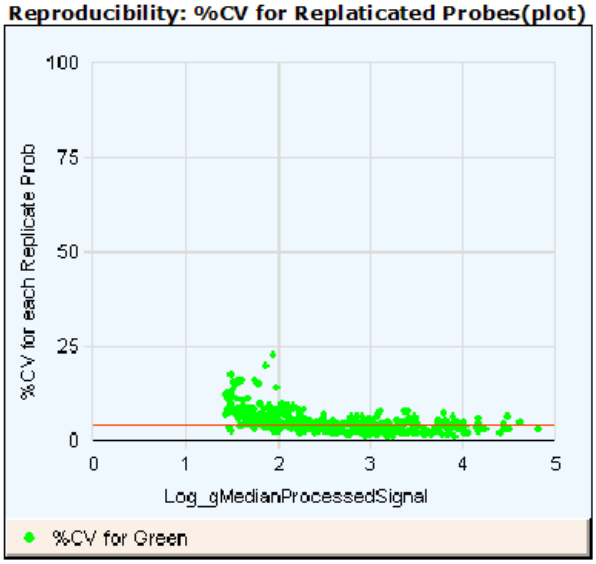


Figure 32 1-color QC Report—Agilent SpikeIns: %CV of Avg. Processed Signal Plot

Reproducibility plot for miRNA (non-control probes)

This graph plots %CV vs. the log_gMedianProcessedSignal for the 1-color miRNA microarray experiment. The region where the %CV flattens out and is not tightly correlated with signal is the range where noise is proportional to signal. This is generally the range used to calculate the median %CV.



Spike-in Signal Statistics

2-color gene expression spike-in signal statistics

These signal statistics and S/N values for spike-ins indicate accuracy and reproducibility of the signals of the microarray probes. The table shows the expected signal of the spike-in probe, the observed average signal, the SD of the observed signal and the S/N of the observed signal.

Agilent SpikeIns Signal Statistics

Probe Name	Exp	Obs	SD	S/N
(+)E1A_r60_n9	-1.00	-0.85	0.02	35.97
(+)E1A_r60_a107	-0.48	-0.38	0.02	18.16
(+)E1A_r60_a135	-0.48	-0.39	0.03	15.02
(+)E1A_r60_n11	-0.48	-0.35	0.03	14.01
(+)E1A_r60_1	0.00	-0.08	0.02	3.77
(+)E1A_r60_a20	0.00	-0.09	0.02	4.16
(+)E1A_r60_3	0.48	0.44	0.02	18.77
(+)E1A_r60_a104	0.48	0.49	0.02	23.07
(+)E1A_r60_a97	0.48	0.56	0.03	21.97
(+)E1A_r60_a22	1.00	0.94	0.02	39.98

Figure 33 QC Report—Agilent SpikeIns Signal Statistics

1-color gene expression spike-in signal statistics

For each sequence of spike-ins this table shows the Probe Name, the median Processed Signal (median of LogProcessedSignal), %CV (SD_ProcessedSignals/Avg_ProcessedSignals) and StdDev (of LogProcessedSignals).

Agilent SpikeIns Signal Statistics

Probe Name	Log(Relative Conc.)	Log(Median Proc. Sig.)	% CV	StdDev
r60_3	0.30	1.01	157.77	0.35
r60_a104	1.30	1.00	88.50	0.27
r60_a107	2.30	1.04	113.17	0.28
r60_a135	3.30	1.72	19.26	0.08
r60_a20	3.83	2.09	7.76	0.03
r60_a22	4.30	2.77	7.32	0.03
r60_a97	4.82	3.17	12.72	0.06
r60_n11	5.30	3.69	8.65	0.04
r60_n9	5.82	4.37	7.19	0.03
r60_1	6.30	4.82	4.85	0.02

Figure 34 1-color QC Report—Agilent SpikeIns Signal Statistics

Spike-in Linearity Check for 2-color Gene Expression

Using the data calculated for the above table, the observed average log ratio is plotted vs the expected log ratio for each of the spike-in probes. A linear regression analysis is done using these values and the metrics are shown below the plot. A slope of 1, y-intercept of 0 and R^2 of 1 is the ideal of such a linear regression. A slope < 1 may indicate compression, such as having under-corrected for background. The regression coefficient (R^2) reflects reproducibility.

The standard deviation for each data point is shown on the plot by an error bar extending above and below the point.

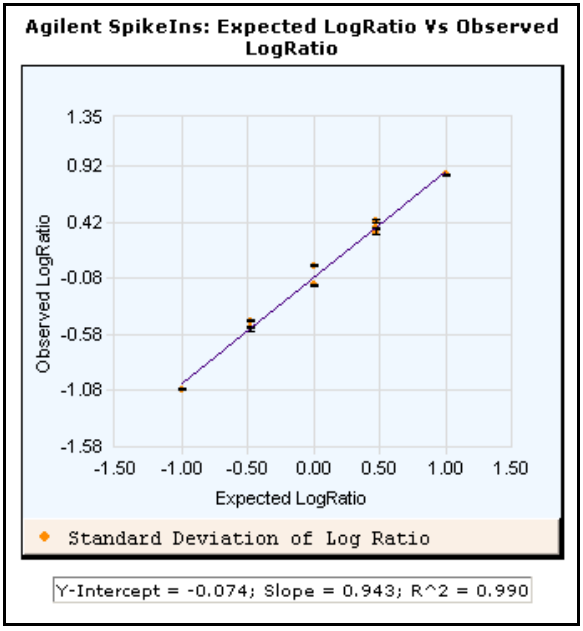


Figure 35 QC Report—Agilent SpikeIns: Expected Log Ratio Vs Observed LogRatio

Spike-in Linearity Check for 1-color Gene Expression

This plot shows the dose/response curve of the spike-ins from the detection limit to the saturation point.

This plot is usually sigmoidal with two asymptotes, one at the scanner saturation point and one at the level of signal for sequences with no specifically bound target. Some microarrays produce plots missing the top asymptote, especially if extended dynamic range is used. (See the plot below.)

At high signal levels the error bars are small since the scanner reaches saturation at this point. Both the signals and standard deviations are underestimated because the saturated data is not excluded from the calculation.

At low signal levels the error bars are visible because the signal is dropping into the background noise. The signal level at the top of the error bars of the features with lowest signal provides a rough estimate of the lower limit of detection. Signals at this level can be slightly overestimated and the error slightly underestimated because the signals below zero are excluded from the calculation.

The most reliable Feature Extraction data is found in the signal range where the signal increases linearly with the concentration of the target.

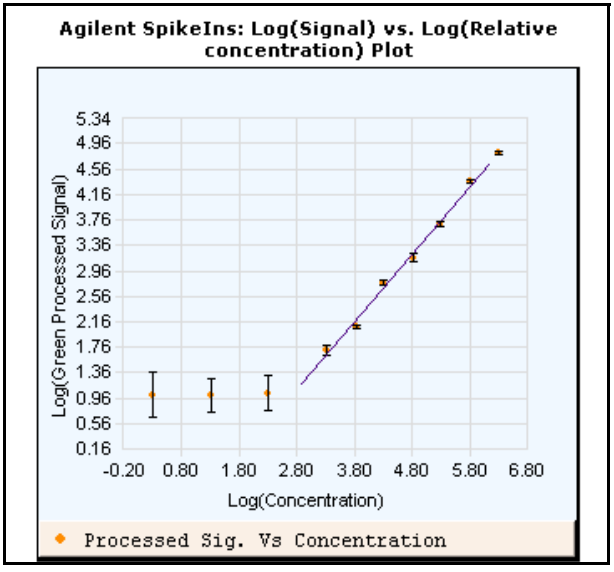


Figure 36 1-color QC Report—Agilent SpikeIns: Log (Signal) vs. Log (Relative concentration) Plot

Table of Values for Concentration-Response Plot (1-color only)

This table presents the values for the log signal vs. log concentration plot shown in Figure 36.

Agilent Spike-In Concentration-Response Statistics

Linear Range Statistics:

Low Signal	1.19
High Signal	4.65
Low Relative Concentration	2.87
High Relative Concentration	6.17
Slope	1.05
R^2 Value	0.99

Detection Limit Statistics

Saturation Point	4.82
Low Threshold	0.97
Low Threshold Error	0.53

Figure 37 1-color QC Report—Agilent Spike-In Concentration-Response Statistics

How the curve and statistics are calculated

Curve fit equation All of the statistics in the table above are calculated using a parameterized sigmoidal curve fit to the data.

$$F(x) = min + \frac{max - min}{1 + e^{(-(x - x0))/w}}$$

where min is the level of signal for sequences with no specifically bound target and max is the upper limit of detection

where $x0$ is the center of the data and close to the center of the linear range

where w is the width of the curve on either side of $x0$.

Curve fit calculations Before the calculations the following assumptions are made:

- Saturation Point is fixed or close to scanner detection limit. This value is Log(Scanner Saturation Value) = 4.82.
- The linear range of the curve, (x0-w) (x0+w), does not define the dynamic range of the data as the data is close to linear for higher multiples of w away from x0.
- The asymptotes for the max and the min are not necessarily symmetric. The upper asymptote is a function of scannner offset, and the lower asymptote is a function of chemistry/scanner noise.

The calculations then follow this order:

a The Min is estimated by taking all the SpikeIn data and for each sequence calculating the BackgroundSubtracted-SignalAverage, the Median of the Log of the processed Signals, StDev of the Log of the processed Signals, the %CV of the processed signals.

The Median Log Proc Signal, %CV, StDev of the Log of the processed signals all show up in the Agilent SpikeIns Signal Statistics table of the QC report.

For each sequence, use the calculated Background-SubtractedSignalAverage and compare against the StdDeviation of the Negative Controls (StdDevBgSubSigNegCtrl) using the formula BGSubAverage * MultErrorGreen > StdDevBgSubSigNegCtrl. Exclude the Proc Signals that fail this test, and use the median of the Proc Signals for the remaining sequences as the initial guess.

b Max is estimated as Log(Scanner SaturationValue).

c x0 is estimated by starting with the y-value (max+min)/2, then finding the 2 closest Med Log Proc Signals above and below this point. Finding the Log(concentrations) of those points and then computing a slope and an intercept by

slope = (MedianLogProcSig[HIGH] – MedianLogProcSig[LOW])/(LogConc[HIGH] – LogConc[LOW]); intercept = LogConc[HIGH] – slope * MedianLogProcSig[HIGH]

d w is estimated by using the slope calculated above. By looking at the derivative of F(x) at x0 we get DF(x):x0 = (max-min)/4*w so w = 4*slope / (max – min).

e After the estimates are complete the data is fit and the parameters (Min,Max, x0, w) are optimized by using a parameterized curve fitting routine (called Levenberg-Marquardt and is a standard technique documented in Numerical Recipes in C on pages 683 – 688).

f After the curve fitting is done, the Low Relative Concentration is calculated as x0 – 2.3*w.

g The High relative Concentration is calculated as x0 + 2.2*w.

h All the eQC points falling between x0 – 2.3*w and x0 + 2.2*w are then fit through a line with the Slope and R-Squared value reported.

i All of the points with a concentration below Low Concentration are used to calculate SpikeIn Detection limit. For each probe, the mean and standard deviation is calculated in linear BGSubSignal space. Then the average plus 1 standard deviation is calculated for each probe. The maximum of these is used. It is converted to log10 space and reported as the SpikeIn Detection Limit.

Relation of curve fit calculations to statistics in table In summary, the table below presents descriptions of the statistics in Figure 37, their definitions within the equation and their output in the stats table.

Table 15 Spike-In Concentration-Response Statistics for 1-color microarrays

Statistic	Description	Where in calculations	Stats Table Output
Saturation Point	upper limit of detection	max–step b	eQCOneColorLogHighSignal
Low Threshold	lower limit of detection	min–step a	eQCOneColorLogLowSignal
Low Threshold Error	error for lower limit	See equation below table	eQCOneColorLogLowSignalError
Low Signal	lowest quantifiable signal in linear range	lowest signal from linear fit in step h	eQCOneColorLinFitLogLowSignal
High Signal	highest quantifiable signal in linear range	highest signal from linear fit in step h	eQCOneColorLinFitLogHighSignal
Low Relative Concentration	lowest concentration leading to quantifiable signal	x0-2.3w in step f	eQCOneColorLinFitLogLowConc
High Relative Concentration	highest concentration leading to quantifiable signal	x0+2.2w in step g	eQCOneColorLinFitLogHighConc
Slope	slope of the linear fit on sigmoidal curve	from step h	eQCOneColorLinFitSlope
R^2 Value	correlation coefficient for linear fit	from step h	eQCOneColorLinFitRSQ
SpikeIn Detection Limit	The average plus 1 standard deviation of the spike ins below the linear concentration range.	from step i	eQCOneColorSpikeInDetectionLimit

$$LowThresholdError = \sqrt{\sum_{A} SD(Log(ProcessedSignals))^2}$$

where the set A is from step a in the table

Accuracy of linear fit to middle of sigmoidal curve Agilent calculated the % difference between expected log processed signals at the high and low relative concentrations on the linear curve with the expected log signals for the same concentrations on the sigmoidal curve.

For the high end of the linear range, the % difference is 15.36%.

For the low end of the linear range, the % difference is 16.75%.

QC Report Results in the FEPARAMS and Stats Tables

See "Parameters/options (FEPARAMS)" on page 113 and "Statistical results (STATS)" on page 133 of this guide for descriptions of the parameters and statistics listed in the tables.

The FEPARAMS table contains most of the QC header information. The Stats table output contains all the metrics shown on the QC Reports. These QC stats let you make "tracking" charts of individual metrics that you may want to follow over time. To separate out the FEPARAMS and Stats tables from each other and the FEATURES table, see "Select to generate a single file for the text output" on page 204 of the *User Guide*.

Table 16 QC Report results present in the text output file (FEPARAMS or STATS tables)*

QC Report Region	Name in QC Report	FEParams or Stats (Green and Red–g(r); Green or Red – no prefix)
"QC Report Headers" on page 72	Date	
	Image	FeatureExtractor_ArrayName
	Protocol	Protocol_Name
	User	FeatureExtractor_UserName
	FE	FeatureExtractor_Version
	Grid	Grid_Name
	BG Method	BGSubtractor_BGSubMethod
For 2-color and CGH, called Spatial Detrend	Background Detrend	BGSubtractor_SpatialDetrendOn
	Multiplicative Detrend	BGSubtractor_MultiplicativeDetrendOn
Not in 1-color	Dye Norm	DyeNorm_CorrMethod
Not in 1-color	Linear DyeNorm Factor	LinearDyeNormFactor
Only green in 1-color	Additive Error	AddErrorEstimateRed & AddErrorEstimateGreen
CGH only	Derivative of Log Ratio Spread	DerivativeofLogRatioSD
XDR only	Saturation Value	g(r)SaturationValue
"Outlier Stats" on page 75	Red Feature NonUniform	NumFeatureNonUnifOL
only green in 1-color	Green Feature NonUniform	NumFeatureNonUnifOL

Table 16 QC Report results present in the text output file (FEPARAMS or STATS tables)*

QC Report Region	Name in QC Report	FEParams or Stats (Green and Red–g(r); Green or Red – no prefix)
	Red Background NonUniform	NumNonUnifBGOL
only green in 1-color	Green Background NonUnif	NumNonUnifBGOL
	Red Feature Population	NumPopnOL
only green in 1-color	Green Feature Population	NumPopnOL
	Red Background Population	NumPopnBGOL
only green in 1-color	Green Background Population	NumPopnBGOL
"Spatial Distribution of All Outliers" on page 75		
only green in 1-color	Red or Green Feat NonUnif	"Calculated on fly"
"Negative Control Stats" on page 78	NetSignal:	
	Average Red	NegCtrlAveNetSig
only green in 1-color	Average Green	NegCtrlAveNetSig
	StdDev Red	NegCtrlSDevNetSig
only green in 1-color	StdDev Green	NegCtrlSDevNetSig
	BG Sub Signal:	
	Average Red	NegCtrlAveBGSubSig
only green in 1-color	Average Green	NegCtrlAveBGSubSig
	StdDev Red	NegCtrlSDevBGSubSig
only green in 1-color	StdDev Green	NegCtrlSDevBGSubSig
"Net Signal Statistics" on page 77	NumSat	CtrleQCNumSatFeat
	99%	CtrleQC99PrcntNetSig
	50%	CtrleQC50PrcntNetSig
	1%	CtrleQC1PrcntNetSig
"Net Signal Statistics" on page 77 (Non-Control Probes)	NumSat	NonCtrlNumSatFeat
	99%	NonCtrl99PrcntNetSig

Table 16 QC Report results present in the text output file (FEPARAMS or STATS tables)*

QC Report Region	Name in QC Report	FEParams or Stats (Green and Red–g(r); Green or Red – no prefix)
	50%	NonCtrl50PrcntNetSig
	1%	NonCtrl1PrcntNetSig
"Net Signal Statistics" on page 77 (Corrected Signals)	No NonControl Feat < 0 Red	NonCtrlNumNegFeatBGSubSig
	No NonControl Feat <0 Green	NonCtrlNumNegFeatBGSubSig
"Local Background Inliers" on page 80	Number Red	LocalBGInlierNum
only green in 1-color	Number Green	LocalBGInlierNum
	Avg Red	LocalBGInlierAve
only green in 1-color	Avg Green	LocalBGInlierAve
	SD Red	LocalBGInlierSDev
only green in 1-color	SD Green	LocalBGInlierSDev
"Foreground Surface Fit" on page 81	RMS_Fit Red	SpatialDetrendRMSFit
only green in 1-color	RMS_Fit Green	SpatialDetrendRMSFit
	RMS_Resid Red	SpatialDetrendRMSFilteredMinusFit
only green in 1-color	RMS_Resid Green	SpatialDetrendRMSFilteredMinusFit
	Avg_Fit Red	SpatialDetrendAveFit
only green in 1-color	Avg_Fit Green	SpatialDetrendAveFit
"Multiplicative Surface Fit" on page 82	RMS_Fit Red	MultDetrendRMSFit
only green in 1-color	RMS_Fit Green	MultDetrendRMSFit
"Spatial Distribution of Up-Regulated and Down-Regulated Features (Positive and Negative Log Ratios)" on page 83	# Up Regulated	NonCtrlNumUpReg
	# Down Regulated	NonCtrlNumDownReg
"Reproducibility Statistics (%CV Replicated Probes)" on page 86	Median %CV red	NonCtrlMedPrctCVBGSubSig
	Median %CV green	NonCtrlMedPrctCVBGSubSig

Table 16 QC Report results present in the text output file (FEPARAMS or STATS tables)*

QC Report Region	Name in QC Report	FEParams or Stats (Green and Red–g(r); Green or Red – no prefix)
"Microarray Uniformity (2-color only)" on page 88	AbsAvgLogRatio Non-Control	NonCtrlAbsAveLogRatio
	AbsAvgLogRatio SpikeIns	eQCAbsAvgLogRatio
	Average S/N Non-Control	NonCtrlSNRLogRatio
	Average S/N SpikeIns	eQCSNRLogRatio
"Sensitivity" on page 89	lowName1	eQCLowSigName1
	lowName2	eQCLowSigName2
	R60_n11 green (lowName1)	eQCSig2BkgLow1
	R60_a97 green (lowName2)	eQCSig2BkgLow2
	Derivative of Log Ratio SD	DerivativeLogRatioSD
"Reproducibility Plots" on page 90	Median %CV red:	eQCMedPrcntCVBGSubSig
only green in 1-color	Median %CV green:	eQCMedPrcntCVBGSubSig
"Spike-in Signal Statistics" on page 93	ProbeName	ProbeName
	Exp	ExpectedLogRatio
	Obs	eQCLogRatioAve
	SD	eQCLogRatioSD
	S/N	eQCLogRatioSNR
"Spike-in Linearity Check for 2-color Gene Expression" on page 95	y-Intercept	eQCObsVsExpLRIntercept
	Slope	eQCObsVsExpLRSlope
	R^2 value	eQCObsVsExpCorr
"Spike-in Signal Statistics" on page 93	Probe Name	ProbeName
	Concentration	GreenStockConc
	Median ObservedProcessed Signal	eQCMedianProcSignal
	Processed Signal %CV	eQCProcSignalPrctCV

Table 16 QC Report results present in the text output file (FEPARAMS or STATS tables)*

QC Report Region	Name in QC Report	FEParams or Stats (Green and Red–g(r); Green or Red – no prefix)
	Processed Signal S/N	eQCProcSignalSNR
"Table of Values for Concentration-Response Plot (1-color only)" on page 97	Saturation Point	eQCOneColorLogHighSignal
	Low Threshold	eQCOneColorLogLowSignal
	Low Threshold Error	eQCOneColorLogLowSignalError
	Low Signal	eQCOneColorLinFitLogLowSignal
	High Signal	eQCOneColorLinFitLogHighSignal
	Low Relative Concentration	eQCOneColorLinFitLogLowConc
	High Relative Concentration	eQCOneColorLinFitLogHighConc
	Slope	eQCOneColorLinFitSlope
	R^2 Value	eQCOneColorLinFitRSQ
	SpikeIn Detection Limit	eQCOneColorSpikeInDetectionLimit

* Results are reported to 9 decimal places in exponential notation for all result files.

QC Metric Set Results

The tables below show the names of the QC metric set results that appear in the Evaluation Tables for each of the three QC metric sets available for Feature Extraction:

- CGH_QCM_Date
- GE1_QCM_Date
- GE2_QCMT_Date

where QCM means QC Metrics and QCMT means QC Metrics with Thresholds.

Table 17 QC metric set results for GE2 Feature Extraction

Name of Metric	FE Stats Used	Description/Measures
AnyColorPrcntFeatNonUnifOL	AnyColorPrcntFeatNonUnifOL	Percentage of Features that are NonUnifOlr in either channel
AnyColorPrcntBGNonUnifOL	AnyColorPrcntBGNonUnifOL	Percentage of LocalBkgdRegions that are NonUnifOlr in either channel
gNegCtrlAveBGSubSig	gNegCtrlAveBGSubSig	Avg of NegControl Bkgd-subtracted signals (Green)
rNegCtrlAveBGSubSig	rNegCtrlAveBGSubSig	Avg of NegControl Bkgd-subtracted signals (Red)
gNegCtrlSDevBGSubSig	gNegCtrlSDevBGSubSig	StDev of NegControl Bkgd-subtracted signals (Green)
rNegCtrlSDevBGSubSig	rNegCtrlSDevBGSubSig	StDev of NegControl Bkgd-subtracted signals (Red)
absE1aObsVsExpSlope	Abs(eQCObsVsExpLRSlope)	Absolute of slope of fit for Observed vs. Expected E1a LogRatios
absE1aObsVsExpCorr	Abs(eQCObsVsExpCorr)	Absolute of correlation of fit for Observed vs. Expected E1a LogRatios
gE1aMedCVBkSubSignal	geQCMedPrcntCVBGSubSig	Median CV of replicated E1a probes: Green Bkgd-subtracted signals
rE1aMedCVBkSubSignal	reQCMedPrcntCVBGSubSig	Median CV of replicated E1a probes: Red Bkgd-subtracted signals

Table 17 QC metric set results for GE2 Feature Extraction

Name of Metric	FE Stats Used	Description/Measures
gNonCntrlMedCVBkSubSignal	gNonCntrlMedCVBkSubSignal	Median CV of replicated NonControl probes: Green Bkgd-subtracted signals
rNonCntrlMedCVBkSubSignal	rNonCntrlMedCVBkSubSignal	Median CV of replicated NonControl probes: Red Bkgd-subtracted signals

Table 18 QC metric set results for GE1 Feature Extraction

Foot note	Name of Metric	FE Stats Used	Description/Measures
	AnyColorPrcntFeatNonUnifOL	AnyColorPrcntFeatNonUnifOL	Percentage of Features that are NonUnifOlr
1	AnyColorPrcntBGNonUnifOL	AnyColorPrcntBGNonUnifOL	Percentage of LocalBkgdRegions that are NonUnifOlr
	gNegCtrlAveBGSubSig	gNegCtrlAveBGSubSig	Avg of NegControl Bkgd-subtracted signals (Green)
	gNegCtrlSDevBGSubSig	gNegCtrlSDevBGSubSig	StDev of NegControl Bkgd-subtracted signals (Green)
	gSpatialDetrendRMSFilteredMinusFit	gSpatialDetrendRMSFilteredMinusFit	Residual of background detrending fit
	gNonCntrlMedCVProcSignal	gMedPrcntCVProcSignal	Median CV of replicated NonControl probes: Green Processed signals (after MultDetrending)
	gE1aMedCVProcSignal	geQCMedPrcntCVProcSignal	Median CV of replicated E1a probes: Green Processed signals (after MultDetrending)
	absGE1E1aSlope	Abs(eQCOneColorLinFitSlope)	Absolute of slope of fit for Signal vs. Concentration of E1a probes
	eQCOneColorLinFitLogLowConc	eQCOneColorLinFitLogLowConc	Log of lowest detectable concentration from fit of Signal vs. Concentration of E1a probes

Table 18 QC metric set results for GE1 Feature Extraction

Foot note	Name of Metric	FE Stats Used	Description/Measures
FOOTNOTE			
1	Data generated before FEv9.5 will not have this metric available in the Stats output. User can make a custom metric which is equivalent: { gNumNonUnifBGOL * 100 / TotalNumFeatures }		

Table 19 QC metric set results for CGH Feature Extraction

Foot note	Name of Metric	FE Stats Used	Description/Measures
1	AnyColorPrcntFeatNonUnifOL	AnyColorPrcntFeatNonUnifOL	Percentage of Features that are NonUnifOlr in either channel
2	DLRSpread	DerivativeOfLogRatioSD	Indicator of noise for baseline of log ratios
2	BGNoiseGreen	gNegCtrlSDevBGSubSig	StDev of NegControl Bkgd-subtracted signals (Green)
2	BGNoiseRed	rNegCtrlSDevBGSubSig	StDev of NegControl Bkgd-subtracted signals (Red)
2	SignalIntensityGreen	gNonCtrl50PrcntBGSubSig	Median Bkgd-subtracted signal of NonControl probes (Green)
2	SignalIntensityRed	rNonCtrl50PrcntBGSubSig	Median Bkgd-subtracted signal of NonControl probes (Red)
2	SignalToNoiseGreen	gNonCtrl50PrcntBGSubSig / gNegCtrlSDevBGSubSig	Ratio of median Signal Intensity to BGNoise (Green)
2	SignalToNoiseRed	NonCtrl50PrcntBGSubSig / rNegCtrlSDevBGSubSig	Ratio of median Signal Intensity to BGNoise (Red)
3	ReproducibilityGreen_BGSubSig	NonCtrlMedPrcntCVBGSubSig / 100	Median CV of replicated NonControl probes: Green Bkgd-subtracted signals
3	ReproducibilityRed_BGSubSig	rNonCtrlMedPrcntCVBGSubSig / 100	Median CV of replicated NonControl probes: Red Bkgd-subtracted signals

Table 19 QC metric set results for CGH Feature Extraction

Foot note	Name of Metric	FE Stats Used	Description/Measures
1	ReproducibilityGreen_ProcSig	gMedPrcntCVProcSignal / 100	Median CV of replicated NonControl probes: Green Processed signals (after MultDetrending)
1	ReproducibilityRed_ProcSig	rMedPrcntCVProcSignal / 100	Median CV of replicated NonControl probes: Red Processed signals (after MultDetrending)

FOOTNOTES

1 New metric, not currently in CGH_Analytics

2 Same metric as currently in CGH_Analytics

3 Same metric calculation, but new name, as currently in CGH_Analytics

3 Text File Parameters and Results

Feature Extraction produces a tab-delimited text file that contains three tables of input parameters and output results.

These tables are FEPARAMS, STATS, and FEATURES. These three tables list all the possible parameters, statistics and feature results that can be generated in the text output file.

FEPARAMS table Contains input parameters and options used to run Feature Extraction.

STATS table Gives results derived from statistical calculations that apply to all features on the microarray.

FEATURES table Displays results for each feature in over 90 output columns, such as gene name, log ratio, processed signal, mean signal, or dye-normalized signal.

You have the option in the Project Properties sheet of selecting to generate either the FULL set of parameters, statistics and feature information, or a COMPACT output package.

The COMPACT output package contains only those columns that are required by GeneSpring, CGH Analytics and Chip Analytics software. The tables on the following pages present both the FULL text files and the COMPACT files.

NOTE Some of the parameters, statistical results and feature results may not be included from any one output file, depending on the protocol used for Feature Extraction.

You also have the option of generating one file with all three tables or three separate files, one for each table. To select to generate one file or three, see "Select to generate a single file for the text output" on page 204 of the User Guide.

To view the text results file in an easy-to-read format, see "View the text result file in Microsoft Excel" on page 87 of the User Guide.

Parameters/options (FEPARAMS)

The top-most section of the result file contains the parameters and option choices that you used to run Feature Extraction.

FULL FEPARAMS Table

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
	Protocol _Name	text	Name of protocol used
	Protocol_date	text	Date the protocol was last modified
	Scan_date	text	Date the image was scanned
	Scan_ScannerName	text	Serial number of the scanner used
	Scan_NumChannels	integer	Number of channels in the scan image
	Scan_MicronsPerPixelX	float	Number of microns per pixel in the X axis of the scan image
	Scan_MicronsPerPixelY	float	Number of microns per pixel in the Y axis of the scan image
	Scan_OriginalGUID	text	The global unique identifier for the scan image
	Grid_Name	text	Grid template name or grid file name
	Grid_Date	integer	Date the grid template or grid file was created
	Grid_NumSubGridRows	integer	Number of subgrid columns
	Grid_NumSubGridCols	integer	Number of subgrid columns
	Grid_NumRows	integer	Number of spots per row of each subgrid
	Grid_NumCols	integer	Number of spots per column of each subgrid
	Grid_RowSpacing	float	Space between rows on the grid

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
	Grid_ColSpacing	float	Space between column on the grid
	Grid_OffsetX	float	In a dense pack array, the offset in the X direction
	Grid_OffsetY	float	In a dense pack array, the offset in the Y direction
	Grid_NomSpotWidth	float	Nominal width in microns of a spot from grid
	Grid_NomSpotHeight	float	Nominal height in microns of a spot from grid
	Grid_GenomicBuild	text	The build of the genome used to create the annotation (if available). If the genome build is not available (not all designs have this information), then it is not put out. All recent and all future designs have it.
	FeatureExtractor_Barcode	text	Barcode of the Agilent microarray read from the scan image
	FeatureExtractor_ScanFileName	text	Name of the scan file used for Feature Extraction
	FeatureExtractor_ArrayName	text	Microarray filename
	FeatureExtractor_DesignFileName	text	Design or grid file used for FE
	FeatureExtractor_PrintingFileName	text	Print file (if available) used for FE
	FeatureExtractor_PatternName	text	Agilent pattern file name
	FeatureExtractor_Extraction Time	text	Time stamp at the beginning of Feature Extraction
	FeatureExtractor_UserName	text	Windows Log-In Name of the User who ran Feature Extraction
	FeatureExtractor_ComputerName	text	Computer name on which Feature Extraction was run
	FeatureExtractor_ScanFileGUID	text	GUID of the scan file

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
	FeatureExtractor_IsXDRExtraction	integer 1 = True 0 = False	Indicates whether or not the extraction was an XDR extraction.
	Scan_NumScanPass	1 or 2	For 5 micron scans , indicates whether the scan mode was a single (1) or double-pass scan mode on the Agilent Scanner.
Place Grid	GridPlacement_Version	text	Version of the grid placement algorithm
Place Grid	GridPlacement_ArrayFormat	integer	Choices for grid placement based on the format of the image. Choices include: Automatically Determine Single Density (11k, 22k) Double Density (44k) 95k 185 (5 and 10 uM) 244 (5 and 10 uM) 25k
Place Grid	GridPlacement_placementMode	integer 0 1	Mode of grid placement Allow the grid to distort Place the grid rigidly allowing only translation and rotation
Place Grid	GridPlacement_ptsPerSide	integer	
Place Grid	GridPlacement_minPeakSpacing	float	
Place Grid	GridPlacement_blocksizeAllowance	float	
Place Grid	GridPlacement_initialFilterWidth	float	
Place Grid	GridPlacement_coarseFilterWidth	float	
Place Grid	GridPlacement_fineFilterWidth	float	
Place Grid	GridPlacement_lineHalfWidth	float	
Place Grid	GridPlacement_halfDepth	float	

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Place Grid	GridPlacement_maximumSlope	float	
Place Grid	GridPlacement_gridWithChannel	integer	
Place Grid	GridPlacement_relativeStartProjection Row	float	
Place Grid	GridPlacement_relativeSizeProjection Row	float	
Place Grid	GridPlacement_relativeStartProjection Col	float	
Place Grid	GridPlacement_relativeSizeProjection Col	float	
Find Spots	SpotAnalysis_Version	text	Version of the spot analysis algorithm
Find Spots	SpotAnalysis_weakthresh	float	Minimum difference between the average intensities of feature and background after Kmeans Initialization
Find Spots	SpotAnalysis_MinimumNumPixels	integer	Minimum number of pixels required for the spot analysis
Find Spots	SpotAnalysis_RegionOfInterest Multiplier	float	Multiplier that defines how big the Region of Interest (ROI) is in terms of nominal spot spacing
Find Spots	SpotAnalysis_convergence_factor	float	Convergence factor of KMeans algorithm
Find Spots	SpotAnalysis_max_em_iter	integer	Maximum number of iterations of the Bayesian Classification
Find Spots	SpotAnalysis_max_reject_ratio	float	Maximum fraction of pixels to be rejected while software performs spotfinding
Find Spots	SpotAnalysis_kmeans_rad_reject_ factor	float	Factor that defines how much individual spot size may vary relative to the nominal spot size

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Find Spots	SpotAnalysis_kmeans_cen_reject_factor	float	Factor that defines how far the actual centroid may move relative to its nominal grid position (in terms of nominal radius). In the protocol this parameter is called the Spot Deviation Limit.
Find Spots	SpotAnalysis_kmeans_moi_reject_factor	float	Maximum allowable moment of inertia of the spot
Find Spots	SpotAnalysis_isspot_factor	float	Factor from the statistics of the found feature and background that indicates if the spot is a spot.
Find Spots	SpotAnalysis_isweakspot_factor	float	Factor from the statistics of the found feature and background that indicates if the spot is a strong one.
Find Spots	SpotAnalysis_BackgroundThreshold	float	Factor by which the individual spot background may vary from the running average of all the background means.
Find Spots	SpotAnalysis_ROIType	integer	Type of Region of Interest
Find Spots	SpotAnalysis_UseNominalDiameterFromGT	integer 1 = True 0 = False	If True, the nominal spot diameter from the grid template is used as a starting point for final spot diameter computation. If False, the nominal diameter is obtained from the grid placement algorithm.
Find Spots	SpotAnalysis_RejectMethod	integer 0 2 3	 Pixel Outlier Rejection turned off Standard Deviation based Interquartile Range based
Find Spots	SpotAnalysis_StatBoundFeat	float	Multiplier parameters for feature outlier rejection method as selected above
Find Spots	SpotAnalysis_StatBoundBG	float	Multiplier parameters for background outlier rejection method as selected above

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Find Spots	SpotAnalysis_SpotStatsMethod	integer	Different algorithms to calculate spot statistics
		1	CookieCutter method
		2	Whole Spot method
Find Spots	SpotAnalysis_CookiePercentage	float	The fraction of the nominal radius used to draw the cookie around the centroid of each spot
Find Spots	SpotAnalysis_ExclusionZone Percentage	float	The outer radius of the exclusion zone based on nominal spot size
Find Spots	SpotAnalysis_EstimateLocalRadius	integer 1 = True 0 = False	The option to calculate the outer radius of the local background based on row and column spacing
Find Spots	SpotAnalysis_LocalBGRadius	float	The outer radius of the local background supplied from the protocol if EstimateLocalRadius is not selected
Find Spots	SpotAnalysis_SignalMethod	integer	The option for the statistical method for determining signals from features: either mean (and standard deviation) or median (and normalized IQR). Mean is 1 and Median is 2.
Flag Outliers	OutlierFlagger_Version	text	Version of Outlier Flagger algorithm
Flag Outliers	OutlierFlagger_NonUnifOLOn	integer	
		1 = True	NonUniformity Outlier flagging turned on
		0 = False	NonUniformity Outlier flagging turned off
Flag Outliers	OutlierFlagger_FeatATerm	float	Applies to feature: specifies the intensity dependent variance and is set to the square of the CV
Flag Outliers	OutlierFlagger_FeatBTerm	float	Applies to feature: specifies the variance due to the Poisson distributed noise

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Flag Outliers	OutlierFlagger_FeatCTerm	float	Applies to feature: specifies variance due to background noise of the scanner, slide glass, and other signal-independent sources
Flag Outliers	OutlierFlagger_BGATerm	float	Applies to background: specifies the intensity-dependent variance and is set to the square of the CV
Flag Outliers	OutlierFlagger_BGBTerm	float	Applies to background: specifies the variance due to the Poisson distributed noise
Flag Outliers	OutlierFlagger_BGCTerm	float	Applies to background: specifies variance due to background noise of the scanner, slide glass, and other signal-independent sources
Flag Outliers	OutlierFlagger_OLAutoComputeABC	integer	
		1 = True	AutoCompute Outlier flagging turned on
		0 = False	AutoCompute Outlier flagging turned off
			For Agilent protocols when this flag is turned on, the polynomial is calculated automatically. This means that all above Feature and BG terms for B and C no longer appear in the output. Rather, they are calculated automatically and appear in the STATS table. Also, the eight parameters following this row appear. For Axon protocols, this flag is always turned on, resulting in new parameters for the output (starting with FeatBCoeff)
Flag Outliers	OutlierFlagger_FeatBCoeff	float	Feature: Red Poissonian Noise Term Multiplier
Flag Outliers	OutlierFlagger_FeatCCoeff	float	Feature: Red Signal Constant Term Multiplier

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Flag Outliers	OutlierFlagger_FeatBCoeff2	float	Feature: Green Poissonian Noise Term Multiplier
Flag Outliers	OutlierFlagger_FeatCCoeff2	float	Feature: Green Signal Constant Term Multiplier
Flag Outliers	OutlierFlagger_BGBCoeff	float	Background: Red Poissonian Noise Term Multiplier
Flag Outliers	OutlierFlagger_BGCCoeff	float	Background: Red Signal Constant Term Multiplier
Flag Outliers	OutlierFlagger_BGBCoeff2	float	Background: Green Poissonian Noise Term Multiplier
Flag Outliers	OutlierFlagger_BGCCoeff2	float	Background: Green Signal Constant Term Multiplier
Flag Outliers	OutlierFlagger_PopnOLOn	integer 1 = True 0 = False	 Population Outlier flagging turned on Population Outlier flagging turned off
Flag Outliers	OutlierFlagger_MinPopulation	integer	Minimum number of replicates to turn on population outlier flagging
Flag Outliers	OutlierFlagger_IQRatio	float	The boundary conditions for conducting box-plot analysis to isolate population outliers
Flag Outliers	OutlierFlagger_BackgroundIQRatio	float	The boundary conditions for conducting box-plot analysis to isolate population outliers for the background
Flag Outliers	OutlierFlagger_Use Qtest	integer 1 = True 0 = False	Enables Qtest statistics when the minimum number of replicates for population outliers is greater than 2 and less than the minimum population specified in the outlier section of the protocol.
Compute Bkgd, Bias and Error	BGSubtractor_MultiplicativeDetrend On	integer 1 = True 0 = False	Enables multiplicative detrending. 1-color and CGH microarray protocols have this parameter enabled.

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Compute Bkgd, Bias and Error	BGSubtractor_MultDetrendWinFilter	integer 0 1 2	No filtering Average filtering Median filtering
Compute Bkgd, Bias and Error	BGSubtractor_MultDetrendIncrement	integer	The increment in number of features by which the square window is shifted horizontally and vertically on the microarray.
Compute Bkgd, Bias and Error	BGSubtractor_MultDetrendWindow	integer	Specifies size of the square window by the number of rows and columns. The specified percentage of low intensity features is selected from this window size.
Compute Bkgd, Bias and Error	BGSubtractor_MultDetrendNeighbor-hoodSize	float [0-1]	Specifies the fraction of total number of neighborhood data points that will be weighted for linear regression during surface fitting for each data point
Compute Bkgd, Bias and Error	BGSubtractor_MultHighPassFilter	integer 1 = True 0 = False	Enables rejection of probes close to zero signal from the set of features used in the fit.
Compute Bkgd, Bias and Error	BGSubtractor_PolynomialMultipli-cativeDetrend	integer 1 = True 0 = False	The option to use a polynomial surface fit method for the multiplicative detrending fit (rather than LOESS).
Compute Bkgd, Bias and Error	BGSubtractor_NegCtrlThresholdMult DetrendFactor	float	This factor multiplies the negative control spread to determine the threshold signal below which low intensity features are filtered out of the multiplicative detrending fit set.
Compute Bkgd, Bias and Error	BGSubtractor_PolynomialMulti-plicativeDetrendDegree	integer [-1, 5]	Shows the degree of the polynomial fit used for the multiplicative detrending. The most common choices are 2 (quadratic or 2nd order surface) and 4 (4th order surface).

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Compute Bkgd, Bias and Error	BGSubtractor_TestMultDetrendOnCVs	integer	Tests whether the replicate CVs improve (i.e. decrease) after multiplicative detrending. If this choice is 1=True, and the replicate CVs don't improve FE doesn't use the multiplicative detrending for that array.
Compute Bkgd, Bias and Error	BGSubtractor_MultDetrendOn Replicates	integer 1 = True 0 = False	Specifies to use only replicated probes (with multiple features) normalized to their replicate average for the multiplicative detrending set.
Compute Bkgd, Bias and Error	BGSubtractor_BGSubMethod	integer	
		1	Either minimum feature or minimum local background across the microarray for background subtraction (global method)
		2	Average of local backgrounds for background subtraction (global method)
		3	Average of negative controls for background for background subtraction (global method)
		5	Local background corresponding to each feature for background subtraction (local method)
		6	Minimum feature across the microarray for background subtraction (global method)
		7	No background subtraction
Compute Bkgd, Bias and Error	BGSubtractor_MaxPVal	float	The pValue at which a feature is determined to be statistically significant above background

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Compute Bkgd, Bias and Error	BGSubtractor_WellAboveMulti	float	The number of standard deviations above background at which the feature is flagged as well above background
Compute Bkgd, Bias and Error	BGSubtractor_BackgroundCorrectionOn	integer	
		1 = True	Globally adjust background turned on
		0 = False	Globally adjust background turned off
Compute Bkgd, Bias and Error	BGSubtractor_BgCorrectionOffset		Adjust the signal of all features by an offset constant so that very low signal features end up at this offset. Appears when Globally adjust background is turned on.
Compute Bkgd, Bias and Error	BGSubtractor_CalculateSurfaceMetricsOn	integer	
		1 = True	Surface fit is done and metrics calculated.
		0 = False	Surface fit and metrics are not done.
Compute Bkgd, Bias and Error	BGSubtractor_SpatialDetrendOn	integer	
		1 = True	Spatial detrend turned on
		0 = False	Spatial detrend turned off
Compute Bkgd, Bias and Error	BGSubtractor_DetrendLowPassFilter	integer	
		1 = True	Low pass filter used
		0 = False	Low pass filter not used
Compute Bkgd, Bias and Error	BGSubtractor_DetrendLowPassPercentage	integer	Specifies percentage of features based on the lowest intensity probes in each window that will be used to fit the surface
Compute Bkgd, Bias and Error	BGSubtractor_DetrendLowPassWindow	integer	Specifies size of the square window by the number of rows and columns. The specified percentage of low intensity features is selected from this window size.
Compute Bkgd, Bias and Error	BGSubtractor_DetrendLowPassIncrement	integer	The increment in number of features by which the above window is shifted horizontally and vertically on the microarray

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Compute Bkgd, Bias and Error	BGSubtractor_NegCtrlSpreadCoeff	float	The number of multiples of the negative control spread that defines the signal range within which features are considered to be within the negative control range for "FeaturesInNegativeControlRange" background detrend option.
Compute Bkgd, Bias and Error	BGSubtractor_NegCtrlSpreadRobustOn	float	Specifies to remove negative control features that are outliers before calculating the negative control spread for use with FeaturesInNegativeControlRange.
Compute Bkgd, Bias and Error	BGSubtractor_AdditiveDetrendFeatureSet	integer 0 1 2	Determines which features are considered for the surface fit set All inlier features Negative control inliers only Features in negative control range
Compute Bkgd, Bias and Error	BGSubtractor_DetrendNeighborhoodSize	float	Specifies the fraction of total number of neighborhood data points that will be weighted for linear regression during surface fitting for each data point
Compute Bkgd, Bias and Error	BGSubtractor_ErrModelSignificance	integer 0 = pixel statistics 1 = error model	Decides whether the error model or pixel staistics are used to determine Positive and Significance calls and WellAboveBackground.

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Compute Bkgd, Bias and Error	BGSubtractor_RobustNCStats	integer 1 = True 0 = False	Specifies if a varation in the population algorithm is turned on. This algorithm repeats the population outlier IQR algorithm on all features classified as negative controls, after the first pass of population algorithm has been run on each sequence. You may want to use this algorithm when you see "hot" features that have not been flagged as population outliers or "hot" sequences where all features of the sequence have higher signals than those in other negative control sequences.
Compute Bkgd, Bias and Error	BGSubtractor_RobustNCOutlierFactor	float	To calculate robust IQR statistics, the algorithm uses upper and lower limits that contain a (Multiplier x IQR) term. This parameter is the Multiplier.
Compute Bkgd, Bias and Error	BGSubtractor_ErrorModel	integer 2 0	Choose universal error, or the most conservative Universal Error Model Most Conservative
Compute Bkgd, Bias and Error	BGSubtractor_MultErrorGreen	float	Multiplicative error component in Green channel
Compute Bkgd, Bias and Error	BGSubtractor_MultErrorRed	float	Multiplicative error component in Red channel
Compute Bkgd, Bias and Error	BGSubtractor_AutoEstimateAddErrorGreen	integer 1 = True 0 = False	 Auto-estimation turned on Auto-estimation turned off
Compute Bkgd, Bias and Error	BGSubtractor_AutoEstimateAddErrorRed	integer 1 = True 0 = False	 Auto-estimation turned on Auto-estimation turned off

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Compute Bkgd, Bias and Error	BGSubtractor_AddErrorGreen	float	This additive error component in the green channel is entered in the protocol when auto-estimation is turned off. When auto-estimation is turned on, the estimated error value appears in the Stats table as AddErrorEstimateGreen.
Compute Bkgd, Bias and Error	BGSubtractor_AddErrorRed	float	This additive error component in the red channel is entered in the protocol when auto-estimation is turned off. When auto-estimation is turned on, the estimated error value appears in the Stats table as AddErrorEstimateRed.
Compute Bkgd, Bias and Error	BGSubtractor_MultNcAutoEstimate	float [0-10]	Multiplier for the first term (standard deviation of the inlier negative control) in the additive error equation.
Compute Bkgd, Bias and Error	BGSubtractor_MultRMSAutoEstimate	float [0-10]	Multiplier for the second term (gMultSpatialDetrendRMSFit) in the additive error equation.
Compute Bkgd, Bias and Error	BGSubtractor_MultResidualsRMSAutoEstimate	float [0-10]	Multiplier for the third term in the additive error equation.
Compute Bkgd, Bias and Error	BGSubtractor_AutoEstimateNCOnly Thresh	float	This parameter is for single density 8-pack microarrays where FE may not be able to accurately subtract the background using the spatial detrending method. This parameter provides a minimum number of features needed for the software to use the residual or the RMS to estimate the additive error. It comes up only if using low density 8-pack microarrays.
Compute Bkgd, Bias and Error	BGSubtractor_UseSurrogates	integer 1 = True 0 = False	Flag indicating the use of surrogates Use of surrogates turned on Use of surrogates turned off
Compute Bkgd, Bias and Error	BGSubtractor_Version	text	Version of BGSubtractor algorithm

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Correct Dye Biases	DyeNorm_Version	text	Version of DyeNorm algorithm
Correct Dye Biases	DyeNorm_SelectMethod	integer	Method for selecting features used for measurement of dye bias:
		4	Use All Probes
		5	Use List of Normalization Genes
		6	Use Rank Consistent Probes
		7	Use Rank Consistent List of Normalization Genes
Correct Dye Biases	DyeNorm_ArePosNegCtrlsOK	integer	
		1 = True	Use positive and negative controls for dye normalization.
		0 = False	Do not use these controls.
Correct Dye Biases	DyeNorm_SignalCharacteristics	integer	
		1	Only positive and significant signals
		2	All positive signals
		3	All negative and positive signals
Correct Dye Biases	DyeNorm_CorrMethod	integer	Methods for computation of dye normalization factor to remove dye bias
		0	Linear
		1	Linear&LOWESS (locally weighted linear regression preceded by linear scaling in each dye channel)
		2	LOWESS (locally weighted linear regression)
Correct Dye Biases	DyeNorm_LOWESSSmoothFactor	float	Smoothing parameter (Neighborhood size) for LOWESS curve fitting
Correct Dye Biases	DyeNorm_LOWESSNumSteps	integer	Number of iterations in LOWESS
Correct Dye Biases	DyeNorm_RankTolerance	float	The threshold to pick rank consistent features between 2 channels for measuring dye biases

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Correct Dye Biases	DyeNorm_VariableRankTolerance	nteger 1 = True 0 = False	Allows the rank tolerance to vary with signal level to allow a fixed percentage of the data to be considered rank consistent.
Correct Dye Biases	DyeNorm_MaxRankedSize	integer	The limit on the number of points used for the dye normalization set. If the number is greater than this, a random subset is chosen using this number of points.
Correct Dye Biases	DyeNorm_IsBGPopnOLOn	integer 1 = True 0 = False	 Software excludes any features from the dye normalization set if the local backgrounds associated with those features have been flagged as population outliers (in either channel). The default recommendation is False.
Compute Ratios	Ratio_Version	text	Version of Ratio algorithm
Compute Ratios	Ratio_PegLogRatioValue	float	Both positive and negative log ratio values are capped to this absolute value
Calculate Metrics	QCMetrics_UseSpikeIns	integer 1 = True 0 = False	 Use SpikeIns Do not use SpikeIns
Calculate Metrics	QCMetrics_minReplicatePopulation	integer	Minimum number of replicates necessary to calculate replicate statistics
Calculate Metrics	QCMetrics_differentialExpression PValue	float	The pValue to use to look for differentially expressed genes
Calculate Metrics	QCMetrics_MaxEdgeDefect Threshold	float	Maximum allowable fraction of features along any edge of the microarray that are non-uniform before a grid placement warning is given.

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
Calculate Metrics	QCMetrics_MaxEdgeNotFound Threshold	float	Maximum allowable fraction of features along any edge of the microarray that are not found before a grid placement warning is given.
Calculate Metrics	QCMetrics_MaxLocalBGNonUnif Threshold	float	Maximum allowable fraction of the local background regions on the microarray that are flagged as NonUniform before a grid placement warning is given.
Calculate Metrics	QCMetrics_MinNegCtrlSDev	float	Minimum value for the standard deviation for the negative controls
Calculate Metrics	QCMetrics_MinReproducibility	float	Minimum value for the reproducibility
Calculate Metrics	QCMetrics_PercentileValuefor Signal	float	The PercentileIntensitySignal is calculated by the software on the [r,g]ProcessedSignal showing the signal at a given percentile over the NonControl features. This parameter is the percentile used for the calculation. By default the value is set to 75; the software generates the 75% Signal value of the ProcessedSignals for all channels available.
	FeatureExtractor_Version	text	Version of Feature Extractor
	FeatureExtractor_SingleTextFile Output	integer 1 = True 0 = False	 The system prints the three tables (FEParams, Stats and Features) are printed in the same text file. The system prints each of the three tables in separate text files.
	FeatureExtractor_JPEGDownSample Factor	float	Factor by which the image is scaled down and then converted to the JPEG format
	FeatureExtractor_ColorMode	integer 0 1	A flag to indicate output color One color; green only 2-color

Table 20 List of parameters and options contained within the FULL text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
	FeatureExtractor_QCReportType	integer 0 1	Type of QC report to generate Gene Expression CGH
	FeatureExtractor_OutputQCReport GraphText	integer 1 = True 0 = False	Generate output details on QC report graphs

COMPACT FEPARAMS Table

Table 21 List of parameters and options contained within the COMPACT text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
	Protocol _Name	text	Name of protocol used
	Protocol_date	text	Date the protocol was last modified
	Scan_date	text	Date the image was scanned
	Scan_ScannerName	text	Serial number of the scanner used
	Scan_NumChannels	integer	Number of channels in the scan image
	Scan_MicronsPerPixelX	float	Number of microns per pixel in the X axis of the scan image
	Scan_MicronsPerPixelY	float	Number of microns per pixel in the Y axis of the scan image
	Scan_OriginalGUID	text	The global unique identifier for the scan image
	Grid_Name	text	Grid template name or grid file name
	Grid_Date	integer	Date the grid template or grid file was created
	Grid_NumSubGridRows	integer	Number of subgrid columns
	Grid_NumSubGridCols	integer	Number of subgrid columns

Table 21 List of parameters and options contained within the COMPACT text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
	Grid_NumRows	integer	Number of spots per row of each subgrid
	Grid_NumCols	integer	Number of spots per column of each subgrid
	Grid_RowSpacing	float	Space between rows on the grid
	Grid_ColSpacing	float	Space between column on the grid
	Grid_OffsetX	float	In a dense pack array, the offset in the X direction
	Grid_OffsetY	float	In a dense pack array, the offset in the Y direction
	Grid_NomSpotWidth	float	Nominal width in microns of a spot from grid
	Grid_NomSpotHeight	float	Nominal height in microns of a spot from grid
	Grid_GenomicBuild	text	The build of the genome used to create the annotation (if available). If the genome build is not available (not all designs have this information), then it is not put out. All recent and all future designs have it.
	FeatureExtractor_Barcode	text	Barcode of the Agilent microarray read from the scan image
	FeatureExtractor_ScanFileName	text	Name of the scan file used for Feature Extraction
	FeatureExtractor_ArrayName	text	Microarray filename
	FeatureExtractor_DesignFileName	text	Design or grid file used for FE
	FeatureExtractor_PrintingFileName	text	Print file (if available) used for FE
	FeatureExtractor_PatternName	text	Agilent pattern file name
	FeatureExtractor_Extraction Time	text	Time stamp at the beginning of Feature Extraction
	FeatureExtractor_UserName	text	Windows Log-In Name of the User who ran Feature Extraction

Table 21 List of parameters and options contained within the COMPACT text output file (FEPARAMS table)

Protocol Step	Parameters	Type/Options	Description
	FeatureExtractor_ComputerName	text	Computer name on which Feature Extraction was run
	FeatureExtractor_ScanFileGUID	text	GUID of the scan file
	FeatureExtractor_Version	text	Version of Feature Extractor
	FeatureExtractor_ColorMode	integer	A flag to indicate output color
		0	One color; green only
		1	2-color
	FeatureExtractor_QCReportType	integer	Type of QC report to generate
		0	Gene Expression
		1	CGH
	FeatureExtractor_IsXDRExtraction	integer	Says if result is from an XDR extraction
		1 = True	
		0 = False	

Statistical results (STATS)

This middle section of the text file describes the results from the statistical calculations. Both the FULL and COMPACT Results are reported to 9 decimal places in exponential notation for all result files.

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gDarkOffsetAverage	rDarkOffsetAverage	float	Average dark offset per image per channel as measured by scanner
gDarkOffsetMedian	rDarkOffsetMedian	float	Median dark offset per image per channel as measured by the scanner
gDarkOffsetStdDev	rDarkOffsetStdDev	float	Standard deviation of the data points measured by the scanner to determine the dark offset per image per channel.
gDarkOffsetNumPts	rDarkOffsetNumPts	integer	Number of points of data measured by the scanner to determine the dark offset per image per channel
gSaturationValue	rSaturationValue	integer	Signal intensity at which spot is considered saturated.
gAvgSig2BkgeQC	rAvgSig2BkgeQC	float	The average ratio of net signal to local background for all spike-in probes
gAvgSig2BkgNegCtrl	rAvgSig2BkgNegCtrl	float	The average ratio of net signal to local background for all negative control probes
gRatioSig2BkgeQC_NegCtrl	rRatioSig2BkgeQC_NegCtrl	float	The ratio of AvgSig2BkgeQC to AvgSig2BkgNegCtrl
gNumSatFeat	rNumSatFeat	integer	The number of saturated features on the microarray per channel
gLocalBGInlierNetAve	rLocalBGInlierNetAve	float	The average of the net signal of all inlier local backgrounds
gLocalBGInlierAve	rLocalBGInlierAve	float	The average of all inlier local backgrounds
gLocalBGInlierSDev	rLocalBGInlierSDev	float	The standard deviation of all inlier local backgrounds

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gLocalBGInlierNum	rLocalBGInlierNum	integer	The number of inlier local backgrounds
gGlobalBGInlierAve	rGlobalBGInlierAve	float	The average of all inliers used in background estimation for the selected global background subtraction method or the average of all inlier local backgrounds if the local background subtraction method is selected (after global background adjustment is applied, if selected)
gGlobalBGInlierSDev	rGlobalBGInlierSDev	float	The standard deviation of all inliers used in background estimation for the selected global background subtraction method or the standard deviation of all inlier local backgrounds if the local background subtraction method is selected
gGlobalBGInlierNum	rGlobalBGInlierNum	integer	The number of all inliers used in background estimation for the selected global background subtraction method or the number of all inlier local backgrounds if the local background subtraction method is selected
gNumFeatureNonUnifOL	rNumFeatureNonUnifOL	integer	The number of features that are flagged as non-uniformity outliers
gNumPopnOL	rNumPopnOL	integer	The number of features that are flagged as population outliers
gNumNonUnifBGOL	rNumNonUnifBGOL	integer	The number of local background regions that are flagged as non-uniformity outliers
gNumPopnBGOL	rNumPopnBGOL	integer	The number of local background regions that are flagged as population outliers
gOffsetUsed	rOffsetUsed	float	Software estimated scanner offset
gGlobalFeatInlierAve	rGlobalFeatInlierAve	float	Average of all inlier features
gGlobalFeatInlierSDev	rGlobalFeatInlierSDev	float	Standard deviation of all inlier features
gGlobalFeatInlierNum	rGlobalFeatInlierNum	float	Number of all inlier features

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
AllColorPrcntSat		float	The percentage of features that are saturated in both the green AND red channels
AnyColorPrcntSat		float	The percentage of features that are saturated in either the green or red channel
AnyColorPrcntFeatNonUnifOL		float	The percentage of features that are feature non-uniformity outliers in either channel
AnyColorPrcntBGNonUnifOL		float	The percentage of local backgrounds that are non-uniformity outliers in either channel
AnyColorPrcntFeatPopnOL		float	The percentage of features that are population outliers in either the green or red channel
AnyColorPrcntBGPopnOL		float	The percentage of local backgrounds that are population outliers in either channel
TotalPrcntFeatOL		float	The percentage of non-control features that are feature non-uniformity outliers in either the green or red channel or are saturated in both channels
gBGAdjust	rBGAdjust	float	Background offset constant to adjust all feature signals. If Adjust Background Globally is set True, all feature signals are adjusted by this offset. If set to the value entered in the protocol, all feature signals are adjusted so that very low level feature signals equal the protocol value.
gNumNegBGSubFeat	rNumNegBGSubFeat	integer	Number of background-subtracted features with negative signals
gNonCtrlNumNegFeatBGSubSig	rNonCtrlNumNegFeatBGSubSig	integer	Number of non-control features with negative background-subtracted signals
gLinearDyeNormFactor	rLinearDyeNormFactor	float	Global dye norm factor

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gRMSLowessDNF	rRMSLowessDNF	float	The root mean square of the average lowess dye norm factor. The lowess dye norm factor for each feature is its DyeNormSignal divided by its BGSubSignal.
DyeNormDimensionlessRMS		float	Dimensionless RMS correction metric (metric that indicates how much correction has been applied based upon the LOWESS curve)
DyeNormUnitWeightedRMS		float	Unit weighted RMS correction metric (metric that indicates how much correction has been applied based upon the LOWESS curve)
gSpatialDetrendRMSFit	rSpatialDetrendRMSFit	float	Root mean square (RMS) of the fitted data points obtained from the Loess algorithm. This gives an idea of the curvature of the surface fit.
gSpatialDetrendRMS Filtered MinusFit	rSpatialDetrendRMS Filtered MinusFit	float	Approximate residual from the surface fit.
gSpatialDetrendSurfaceArea	rSpatialDetrendSurfaceArea	float	Normalized area–the fitted surface area divided by the projected area on the microarray; also gives an idea of the curvature of the surface gradient.
gSpatialDetrendVolume	rSpatialDetrendVolume	float	Sum of the intensities of the surface area minus the offset. The offset is calculated as the volume under the flat surface (parallel to the glass slide) passing through the minimum intensity point of the fitted surface. This number (total volume - offset) is normalized by the area of the microarray.
gSpatialDetrendAveFit	rSpatialDetrendAveFit	float	Describes the average intensity of the surface gradient
gNonCtrlNumSatFeat	rNonCtrlNumSatFeat	integer	The number of saturated non-control features

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gNonCtrl99PrcntNetSig	rNonCtrl99PrcntNetSig	float	NetSignal intensity at 99th percentile for all non-control probes
gNonCtrl50PrcntNetSig	rNonCtrl50PrcntNetSig	float	NetSignal intensity at 50th percentile for all non-control probes
gNonCtrl1PrcntNetSig	rNonCtrl1PrcntNetSig	float	NetSignal intensity at 1st percentile for all non-control probes
gNonCtrlMedPrcntCVBGSubSig	rNonCtrlMedPrcntCVBGSubSig	float	The median percent CV of background-subtracted signals for inlier noncontrol probes
gCtrleQCNumSatFeat	rCtrleQCNumSatFeat	integer	The number of saturated spike-in features
gCtrleQC99PrcntNetSig	rCtrleQC99PrcntNetSig	float	NetSignal intensity at 99th percentile of all spike-in probes
gCtrleQC50PrcntNetSig	rCtrleQC50PrcntNetSig	float	NetSignal intensity at 50th percentile of all spike-in probes
gCtrleQC1PrcntNetSig	rCtrleQC1PrcntNetSig	float	NetSignal intensity at 1st percentile of all spike-in probes
geQCMedPrcntCVBGSubSig	reQCMedPrcntCVBGSubSig	float	The median percent CV of background-subtracted signals for inlier spike-in probes
geQCSig2BkgLow1	reQCSig2BkgLow1	float	Median ratio (net signal to BGUsed) of all inlier features for an spike-in probe with lowest concentration spiked in red and green channels
geQCSig2BkgLow2	reQCSig2BkgLow2	float	Median ratio (net signal to BGUsed) of all inlier features for an spike-in probe with second lowest concentration spiked in red and green channels
gNegCtrlNumInliers	rNegCtrlNumInliers	integer	Number of all inlier negative controls
gNegCtrlAveNetSig	rNegCtrlAveNetSig	float	Average net signal of all inlier negative controls

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gNegCtrlSDevNetSig	rNegCtrlSDevNetSig	float	Standard deviation of the net signal of all inlier negative controls
gNegCtrlAveBGSubSig	rNegCtrlAveBGSubSig	float	Average background-subtracted signal of all inlier negative controls
gNegCtrlSDevBGSubSig	rNegCtrlSDevBGSubSig	float	Standard deviation of the background-subtracted signals of all inlier negative controls
gAveNumPixOLLo	rAveNumPixOLLo	integer	The average number of pixels that are rejected from each feature at the low end of the intensity spectrum
gAveNumPixOLHi	rAveNumPixOLHi	integer	The average number of pixels that are rejected from each feature at the high end of the intensity spectrum
gPixCVofHighSignalFeat	rPixCVofHighSignalFeat	float	Average of pixel CV for features with high signal
gNumHighSignalFeat	rNumHighSignalFeat	integer	The number of features with high signal
NonCtrlAbsAveLogRatio		float	This result is from a two-step calculation. Step 1 for each probe calculates the absolute average log ratio of all inlier non-control features with minimum number of replicates. Step 2 calculates the average of all absolute average log ratios calculated in step 1.
NonCtrlSDevLogRatio		float	The average standard deviation of log ratios of all inlier non-control probe sets with a minimum number of replicates
NonCtrlSNRLogRatio		float	The average of signal to noise values of the log ratio for all inlier non-control probe sets with a minimum number of replicates

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
eQCAbsAveLogRatio		float	This result is from a two-step calculation. Step 1 for each probe calculates the absolute average log ratio of all inlier spikein features with minimum number of replicates. Step 2 calculates the average of all absolute average log ratios calculated in step 1.
eQCSDevLogRatio		float	Average standard deviation of log ratios of all inlier spike-in probe sets with a minimum number of replicates
eQCSNRLogRatio		float	Average signal to noise value of log ratios of all inlier spike-in probe sets with a minimum number of replicates
AddErrorEstimateGreen		float	The additive error estimated for the microarray in the green channel.
AddErrorEstimateRed		float	The additive error estimated for the microarray in the red channel.
TotalNumFeatures		integer	Total number of features that show up in output file.
NonCtrlNumUpReg		integer	Number of up-regulated non-control probes
NonCtrlNumDownReg		integer	Number of down-regulated non-control probes
eQCObsVsExpLRSlope		float	For 2-color QC report: Slope of the linear regression fit of the plot of the expected versus observed average log ratio for each spike-in probe
eQCObsVsExpLRIntercept		float	For 2-color QC report: Intercept of the linear regression fit of the plot of the expected versus observed average log ratio for each spike-in probe

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
eQCObsVsExpCorr		float	For 2-color QC report: The R2 value of the linear regression fit of the plot of the expected versus observed average log ratio for each spike-in probe
NumIsNorm		integer	Number of features used for normalization
ROI Width ROI Height		float	The width or height (in pixels) of the region of interest (ROI) about a nominal spot location. The spotfinder determines the found centroid and spot size of the spot within the ROI.
CentroidDiffX		float	The average absolute of difference between nominal centroids and corresponding found centroids in X direction
CentroidDiffY		float	The average absolute of difference between nominal centroids and corresponding found centroids in Y direction
NumFoundFeat		integer	The number of features that are flagged as found
MaxNonUnifEdges		float	Maximum fraction of features that are non-uniform along any edge of the microarray
MaxSpotNotFoundEdges		float	Maximum fraction of features that are not found along any edge of the microarray
gMultDetrendRMS Fit	rMultDetrendRMS Fit	float	Root mean square (RMS) of the fitted data points obtained from the second degree polynomial equation in Multiplicative Detrending. This gives an idea of the curvature of the surface fit to the "hybridization dome" in the Agilent Hybridization chambers.

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gMultDetrendSurfaceAverage	rMultDetrendSurfaceAverage	float	The average of the surface calculated by multiplicative detrending. This average is used to normalize the surface. It is a straight average over all the points in the surface.
DerivativeOfLogRatioSD		float	Measures the standard deviation of the probe-to-probe difference of the log ratios. This is a metric used in CGH experiments where differences in the log ratios are small on average. A smaller standard deviation here indicates less noise in the biological signals.
eQCLowSigName1		text	The probe name of the eQC probe spiked in at the lowest concentration.
eQCLowSigName2		text	The probe name of the eQC probe spiked in at the second lowest concentration.
eQCOneColorLogLowSignal			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Log of low signal for the data
eQCOneColorLogLowSignal-Error			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Error in the log of low signal for the data
eQCOneColorLogHighSignal			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Log of high signal for the data
eQCOneColorLinFitLogLowConc			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Log of low concentration in the linear range of curve fit
eQCOneColorLinFitLogLow-Signal			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Log of low signal in the linear range of curve fit

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
eQCOneColorLinFitLogHigh-Conc			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Log of high concentration in the linear range of curve fit
eQCOneColorLinFitLogHigh-Signal			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Log of high signal in the linear range of curve fit
eQCOneColorLinFitSlope			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Slope of the linear range of curve fit
eQCOneColorLinFitIntercept			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Intercept of the linear range of curve fit
eQCOneColorLinFitRSQ			Agilent Spike-In Concentration-Response Statistic in the 1-color QC Report: Square of the correlation coefficient of the linear range of curve fit.
eQCOneColorSpikeDetection-Limit			The detection limit as determined by measuring the average plus 1 standard deviation of all spike-in probes below the linear concentration range. This value is the maximum of these.
gNonCtrl50PrcntBGSubSig	gNonCtrl50PrcntBGSubSig	float	Background-subtracted signal intensity at 50th percentile for all non-control probes.
gCtrleQC50PrcntBGSubSig	rCtrleQC50PrcntBGSubSig	float	The median background-subtracted signal for all the embedded QC probes on the microarray.

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gMedPrcntCVProcSignal	rMedPrcntCVProcSignal	float	The median %CV for replicate non-control probes using the processed signal. This value is calculated by calculating the average, SD and %CV of the processed signal of each replicated probe. For non-control replicated probes, there must be at least 10 CVs from which to calculate a median; otherwise, -1 is reported. The MedPrcntCVProcSignal and the MedPrcntCVBGSubSignal show if Multiplicative Detrending is having a positive effect on the data. If multiplicative detrending is helping, the MedPrcntCVProcSignal should be smaller than the MedPrcntCVBGSubSignal.
geQCMedPrcntCVProcSignal	reQCMedPrcntCVProcSignal	float	This is the same as MedPrcntCVProcSignal, except that it is performed using the eQC SpikeIn Replicates rather than the nonControl Replicates. There must be at least 3 CVs from which to calculate a median.
gOutlierFlagger_Auto_FeatB Term	rOutlierFlagger_Auto_FeatB Term	float	Applies to feature: specifies the variance due to the Poisson distributed noise; automatically calculated when OLAutoCompute is turned on
gOutlierFlagger_Auto_FeatC Term	rOutlierFlagger_Auto_FeatC Term	float	Applies to feature: specifies variance due to background noise of the scanner, slide glass, and other signal-independent sources; automatically calculated when OLAutoCompute is turned on
gOutlierFlagger_Auto_BgndB Term	rOutlierFlagger_Auto_BgndB Term	float	Applies to background: specifies the variance due to the Poisson distributed noise; automatically calculated when OLAutoCompute is turned on

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gOutlierFlagger_Auto_BgndC Term	rOutlierFlagger_Auto_BgndC Term	float	Applies to background: specifies variance due to background noise of the scanner, slide glass, and other signal-independent sources; automatically calculated when OLAutoCompute is turned on
OutlierFlagger_FeatChiSq		float	Confidence Interval for the feature
OutlierFlagger_BgndChiSq		float	Confidence Interval for the background
gXDRLowPMTSlope	rXDRLowPMTSlope		The slope that is multiplied by the original low intensity Mean Signal to get the XDR mean signal. Used in the linear equation relating the Mean (or Median) Signal in the low intensity scan to the scaled intensity used in the combined XDR output.
gXDRLowPMTIntercept	rXDRLowPMTIntercept		The intercept that is added to the Slope*LowIntensityMeanSignal to get the XDR Mean Signal. Used in the linear equation relating the Mean (or Median) Signal in the low intensity scan to the scaled intensity used in the combined XDR output.
GriddingStatus		integer	Indicates that the automatic image processing was flagged as needing evaluation.
TotalNumberOfReplicated Genes		integer	Number of genes that have replicate features on the array.
gMultDetrendMeanSignal Difference			This is output for miRNA only. If multiplicative detrending is turned on, the meanSignal over all replicated noncontrols is calculated bfore detrending and after detrending. The difference in mean signals is reported here. Because the mean signal should not change, this number should be close to 0. Without Multiplicative detrending this number is always 0.

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
EffectiveFeatureSizeFraction		float	Estimates the ratio of the effective feature size to the nominal feature size. It is calculated by looking at the ratio of the whole spot measurement versus the cookie measurement.
Feature UniformityAnomaly Fraction		float	Fraction (Num/TotalNum) of the number of features looked at that had anomalous ratios. This gives a measure of the percentage of representative spots that are strange (e.g., donuts, super hot spots, hot crescents).
UsedDefaultEffectiveFeature Size		float	Reports whether or not the default effective feature size was used. If the default was used, the stat is 1. If the effective feature size was estimated, the stat value is 0.
gPercentileIntensityProcessed Signal	rPercentileIntensityProcessed Signal		The protocol lets you enter the Percentile Value at which the intensity of the noncontrol signals is recorded. All protocols specify the 75th percentile. This number is the intensity of all the noncontrol signals in the 75th percentile. This stat is used to normalize 1-color data.
NumGeneNonUnifOL		integer	Number of genes that do not have any replicate features on the array where both color channels are not Feature Non-Uniform outliers. If multiple probes address the same gene, this value actually states the number of probes that have no non-uniform replicates.
gTotalSignal99pctile		float	These are metrics for miRNA only. This is the value of the TotalGeneSignal for all genes at the 99th percentile.
gTotalSignal75pctile		float	These are metrics for miRNA only. This is the value of the TotalGeneSignal for all genes at the 75th percentile.

Table 22 Stats results contained in the text output file (STATS table)*

Stats (Green Channel)	Stats (Red Channel)	Type	Description
gNegCtrlSpread	rNegCtrlSpread	float	The root mean square (RMS) of the preliminary spatial fit of the negative controls. It is equivalent to a standard deviation of NC signals after removal of spatial homogeneities. Used as a preliminary estimation of the noise on the array for selecting near-zero probes in spatial detrending, and conversely for excluding near-zero probes in multiplicative detrending.
gNonCtrlNumWellAboveBG	rNonCtrlNumWellAboveBG		Measure of the number of noncontrol features whose signals are well above background. Used as a metric for the number of features with significant signal.
Metric_*MetricName*			The name of a metric in the metric set. The given value is the one that has been calculated for this metric. You can have more than one metric in a given metric set.
Metric_*MetricName*_IsInRange		Integer: 1=in range; 0=out of range	Indicates whether the metric was within any user-defined thresholds found in the metric set for that metric.
Metric_*MetricName*_Extraction Status		Integer 0=in range; 1=out of range	This is put out only if a metric set has been run. It gives a status of the overall array.
Metric_*MetricName*_QCMetric Results		String	If the Extraction Status = 0, the output says ExtractionInRange. If the Extraction Status = 1, the output says ExtractionEvaluate.

* Results are reported to 9 decimal places in exponential notation for all result files.

Feature results (FEATURES)

The bottom section of the text file gives descriptions of the results for each feature. Results are reported to 9 decimal places in exponential notation for all result files.

FULL Features Table

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
FeatureNum		integer		Feature number
Row		integer		Feature location : row
Col		integer		Feature location : column
SubTypeMask		integer		Numeric code defining the subtype of any control feature
SubTypeName		integer		Name of the subtype of any control feature
Start		integer		Indicates the place in the transcript where the probe sequence starts.
Sequence		text		The sequence of bases printed on the array.
ProbeUID		integer		Unique integer for each unique probe in a design

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
ControlType		integer		Feature control type (See "XML Control Type output" on page 181 for definitions.)
			0	Control type none
			1	Positive control
			-1	Negative control
			-20000	Not probe (See Ch. 4 for definition)
			-30000	Ignore (See Ch. 4 for definition)
ProbeName		text		An Agilent-assigned identifier for the probe synthesized on the microarray
GeneName		text		This is an identifier for the gene for which the probe provides expression information. The target sequence identified by the systematic name is normally a representative or consensus sequence for the gene.
SystematicName		text		This is an identifier for the target sequence that the probe was designed to hybridize with. Where possible, a public database identifier is used (e.g., TAIR locus identifier for Arabidopsis). Systematic name is reported ONLY if Gene name and Systematic name are different.
Description		text		
PositionX PositionY		float		Found coordinates of the feature centroid in microns

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
LogRatio (base 10)		float		log(REDsignal/GREENsignal) per feature (processed signals used)
				If SURROGATES are turned off, then:
			-4	if DyeNormRedSig <= 0.0 & DyeNormGreenSig > 0.0
			4	if DyeNormRedSig > 0.0 & DyeNormGreenSig <= 0.0
			0	if DyeNormRedSig <= 0.0 & DyeNormGreenSig <= 0.0
LogRatioError		float		If SURROGATES are turned off, then:
			1000	if DyeNormRedSig <= 0.0 OR DyeNormGreenSig <= 0.0
				IF SURROGATES are turned on, then:
				LogRatioError = error of the log ratio calculated according to the error model chosen
PValueLogRatio		float		Significance level of the Log Ratio computed for a feature
gSurrogateUsed	rSurrogateUsed	float	Non-zero value	The g(r) surrogate value used
			0	No surrogate value used

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gIsFound	rIsFound	boolean	1 = IsFound 0 = IsNotFound	A boolean used to flag found features. The flag is applied independently in each channel. A feature is considered Found if two conditions are true: 1) the difference between the feature signal and the local background signal is more than 1.5 times the local background noise and 2) the spot diameter is at least 0.30 times the nominal spot diameter.
gProcessedSignal	rProcessedSignal	float		The signal left after all the FE processing steps have been completed. In the case of one color, ProcesssedSignal contains the Multiplicatively Detrended BackgroundSubtracted Signal if the detrending is selected and helps. If the detrending does not help, this column will contain the BackgroundSubtractedSignal.
gProcessedSigError	rProcessedSigError	float		The universal or propagated error left after all the processing steps of Feature Extraction have been completed. In the case of one color, ProcessedSignalError has had the Error Model applied and will contain at least the larger of the universal (UEM) error or the propagated error. If multiplicative detrending is performed, ProcessedSignalError contains the error propagated from detrending. This is done by dividing the error by the normalized MultDetrendSignal.

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gNumPixOLHi	rNumPixOLHi	integer		Number of outlier pixels per feature with intensity > upper threshold set via the pixel outlier rejection method. The number is computed independently in each channel. These pixels are omitted from all subsequent calculations.
gNumPixOLLo	rNumPixOLLo	integer		Number of outlier pixels per feature with intensity < lower threshold set via the pixel outlier rejection method. The number is computed independently in each channel. These pixels are omitted from all subsequent calculations. NOTE: The pixel outlier method is the ONLY step that removes data in Feature Extraction.
gNumPix	rNumPix	integer		Total number of pixels used to compute feature statistics; ie. total number of inlier pixels/per spot; same in both channels
gMeanSignal	rMeanSignal	float		Raw mean signal of feature in green (red) channel (inlier pixels)
gMedianSignal	rMedianSignal	float		Raw median signal of feature in green (red) channel (inlier pixels)
gPixSDev	rPixSDev	float		Standard deviation of all inlier pixels per feature; this is computed independently in each channel.
gPixNormIQR	rPixNormIQR	float		The normalized Inter-quartile range of all of the inlier pixels per feature. The range is computed independently in each channel.
gBGNumPix	rBGNumPix	integer		Total number of pixels used to compute local BG statistics per spot; ie. total number of BG inlier pixels; same in both channels

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gBGMeanSignal	rBGMeanSignal	float		Mean local background signal (local to corresponding feature) computed per channel (inlier pixels)
gBGMedianSignal	rBGMedianSignal	float		Median local background signal (local to corresponding feature) computed per channel (inlier pixels)
gBGPixSDev	rBGPixSDev	float		Standard deviation of all inlier pixels per local BG of each feature, computed independently in each channel
gBGPixNormIQR	rBGPixNormIQR	float		The normalized Inter-quartile range of all of the inlier pixels per local BG of each feature. The range is computed independently in each channel.
gNumSatPix	rNumSatPix	integer		Total number of saturated pixels per feature, computed per channel
gIsSaturated	rIsSaturated	boolean	1 = Saturated or 0 = Not saturated	Boolean flag indicating if a feature is saturated or not. A feature is saturated IF 50% of the pixels in a feature are above the saturation threshold.
gIsLowPMTScaled Up	rIsLowPMTScaled Up	boolean	1 = Low 0 = High	Reports if the feature signal value is from the scaled-up low signal image or from the high signal image
PixCorrelation		float		Ratio of estimated feature covariance in RedGreen space to product of feature standard deviation in Red Green space The covariance of two features measures their tendency to vary together, i.e., to co-vary. In this case, it is a cumulative quantitation of the tendency of pixels belonging to a particular feature in Red and Green spaces to co-vary.
BGPixCorrelation		float		The same concept as above but in case of background.

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gIsFeatNonUnifOL	rIsFeatNonUnifOL	boolean	g(r)IsFeatNonUnifOL = 1 indicates Feature is a non-uniformity outlier in g(r)	Boolean flag indicating if a feature is a NonUniformity Outlier or not. A feature is non-uniform if the pixel noise of feature exceeds a threshold established for a "uniform" feature.
gIsBGNonUnifOL	rIsBGNonUnifOL	boolean	g(r)IsBGNonUnifOL = 1 indicates Local background is a non-uniformity outlier in g(r)	The same concept as above but for background.
gIsFeatPopnOL	rIsFeatPopnOL	boolean	g(r)IsFeatPopnOL = 1 indicates Feature is a population outlier in g(r)	Boolean flag indicating if a feature is a Population Outlier or not. Probes with replicate features on a microarray are examined using population statistics. A feature is a population outlier if its signal is less than a lower threshold or exceeds an upper threshold determined using a multiplier (1.42) times the interquartile range (i.e., IQR) of the population.
gIsBGPopnOL	rIsBGPopnOL	boolean	g(r)IsBGPopnOL = 1 indicates local background is a population outlier in g(r)	The same concept as above but for background
IsManualFlag		boolean		You can enter a boolean to flag features for downstream filtering in third party gene expression software.
gBGSubSignal	rBGSubSignal	float	g(r)BGSubSignal = g(r)MeanSignal - g(r)BGUsed	Background-subtracted signal. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gBGSubSigError	rBGSubSigError	float		Propagated standard error as computed on net g(r) background-subtracted signal. For one color, the error model is applied to the background-subtracted signal. This will contain the larger of he universal (UEM) error or the propagated error.
BGSubSigCorrelation		float		Ratio of estimated background-subtracted feature signal covariance in RG space to product of background-subtracted feature standard deviation in RG space
gIsPosAndSignif	rIsPosAndSignif	Boolean	g(r)isPosAndSignif = 1 indicates Feature is positive and significant above background	Boolean flag, established via a 2-sided t-test, indicates if the mean signal of a feature is greater than the corresponding background (selected by user) and if this difference is significant. To view variables used in the t-test, see Table 33 on page 213.
gPValFeatEqBG	rPValFeatEqBG	float		pValue from t-test of significance between g(r)Mean signal and g(r) background (selected by user)
gNumBGUsed	rNumBGUsed	integer		Number of local background regions or features used to calculate the background used for background subtraction on this feature.
gIsWellAboveBG	rIsWellAboveBG	Boolean		Boolean flag indicating if a feature is WellAbove Background or not, feature passes g(r)IsPosAndSignif and additionally the g(r)BGSubSignal is greater than 2.6*g(r)BG_SD. You can change the multiplier 2.6.

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gBGUsed	rBGUsed	float	g(r)BGSubSignal = g(r)MeanSignal - g(r)BGUsed	Background used to subtract from the MeanSignal; variable also used in t-test. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.
gBGSDUsed	rBGSDUsed	float		Standard deviation of background used in g(r) channel; variable also used in t-test and surrogate algorithms. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, seeTable 33 on page 213.
IsNormalization		boolean	1 = Feature used; 0 = Feature not used	A boolean flag which indicates if a feature is used to measure dye bias
gDyeNormSignal	rDyeNormSignal	float		The dye-normalized signal in the indicated channel
gDyeNormError	rDyeNormError	float		The standard error associated with the dye-normalized signal
DyeNormCorrelation		float		Dye-normalized red and green pixel correlation
ErrorModel			0 = Propagated model chosen by you or by software 1 = Universal error model chosen by you or by software	Indicates the error model that you chose for Feature Extraction or that the software uses if you have chosen the "Most Conservative" option
xDev		float		A signal-to-noise parameter used to calculate pValue; calculated differently depending on error model chosen

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gSpatialDetrendIsIn FilteredSet	rSpatialDetrendIsIn FilteredSet	Boolean	1 = Feature in filtered set 0 = Feature not in filtered set	Set to true for a given feature if it is part of the filtered set used to detrend the background. This feature is considered part of the locally weighted lowest x% of features as defined by the *DetrendLowPassPercentage*.
gSpatialDetrend SurfaceValue	rSpatialDetrend SurfaceValue	float		Value of the smoothed surface calculated by the Spatial detrend algorithm
gIsLowEnoughAdd Detrend	rIsLowEnoughAdd Detrend	boolean		These points are considered to be in the background for the purposes of spatial detrending and multiplicative detrending. If the Boolean value is true for a given point, it will be used in spatial detrending and not in multiplicative detrending (depends on parameters).
SpotExtentX		float		Diameter of the spot (X-axis)
SpotExtentY		float		Diameter of the spot (Y-axis)
gNetSignal	rNetSignal	float		MeanSignal minus DarkOffset
gTotalProbeSignal		float		This signal is the robust average of all the processed green signals for each replicated probe multiplied by the total number of probe replicates, the EffectiveFeature SizeFraction, the Nominal Spot Area and the Weight. For miRNA analyses
gTotalProbeError		float		This error is the robust average of all the processed green signal errors for each replicated probe multiplied by the total number of probe replicates, the EffectiveFeature SizeFraction, the Nominal Spot Area and the Weight. For miRNA analyses
gTotalGeneSignal		float		This signal is the total probe signal times the number of probes per gene. For miRNA analyses

Table 23 Feature results contained in the FULL output text file (FULL FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gTotalGeneError		float		This error is the square root of the square of the (total probe error times the number of probes per gene). For miRNA analyses
gIsGeneDetected		boolean		Lets you know if the gene was detected on the miRNA microarray.
gMultDetrendSignal	rMultDetrendSignal	float		A surface is fitted through the log of the background-subtracted signal to look for multiplicative gradients. A normalized version of that surface interpolated at each point of the microarray is stored in MultDetrendSignal. The surface is normalized by dividing each point by the overall average of the surface. That average is stored in MultDetrendSurfaceAverage as a statistic. 1-color only
gProcessed Background	rProcessed Background	float		Indicates the Background signal that was selected to be used (Mean or Median).
gProcessedBkng Error	rProcessedBkng Error	float		Indicates the Background error that was selected to be used (PixSD or NormIQR)
IsUsedBGAdjust		Boolean	1 = Feature used 0 = Feature not used	A Boolean used to flag features used for computation of global BG offset
gInterpolatedNeg CtrlSub	rInterpolatedNeg CtrlSub	float		Value at the polynomial fit of the negative controls.
gIsInNegCtrlRange	rIsInNegCtrlRange	boolean		Set to true for a given feature if its signal intensity is in the negative control range.
gIsUsedInMD	rIsUsedInMD	boolean		Indicates whether this feature was included in the set used to generate the multiplicative detrend surface.

* Results are reported to 9 decimal places in exponential notation for all result files.

COMPACT Features Table

Table 24 Feature results contained in the COMPACT output text file (COMPACT FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
FeatureNum		integer		Feature number
Row		integer		Feature location : row
Col		integer		Feature location : column
SubTypeMask		integer		Numeric code defining the subtype of any control feature
ControlType		integer		Feature control type (See "XML Control Type output" on page 181 for definitions.)
			0	Control type none
			1	Positive control
			-1	Negative control
			-20000	Not probe (See Ch. 4 for definition)
			-30000	Ignore (See Ch. 4 for definition)
ProbeName		text		An Agilent-assigned identifier for the probe synthesized on the microarray
SystematicName		text		This is an identifier for the target sequence that the probe was designed to hybridize with. Where possible, a public database identifier is used (e.g., TAIR locus identifier for Arabidopsis). Systematic name is reported ONLY if Gene name and Systematic name are different.

Table 24 Feature results contained in the COMPACT output text file (COMPACT FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
LogRatio (base 10)		float		log(REDsignal/GREENsignal) per feature (processed signals used) If SURROGATES are turned off, then: if DyeNormRedSig <= 0.0 & DyeNormGreenSig > 0.0 -4 if DyeNormRedSig > 0.0 & DyeNormGreenSig <= 0.0 4 if DyeNormRedSig <= 0.0 & DyeNormGreenSig <= 0.0 0
LogRatioError		float		If SURROGATES are turned off, then: 1000 if DyeNormRedSig <= 0.0 OR DyeNormGreenSig <= 0.0 IF SURROGATES are turned on, then: LogRatioError = error of the log ratio calculated according to the error model chosen
PValueLogRatio		float		Significance level of the Log Ratio computed for a feature
gProcessedSignal	rProcessedSignal	float		The signal left after all the FE processing steps have been completed. In the case of one color, ProcesssedSignal contains the Multiplicatively Detrended BackgroundSubtracted Signal if the detrending is selected and helps. If the detrending does not help, this column will contain the BackgroundSubtractedSignal.

Table 24 Feature results contained in the COMPACT output text file (COMPACT FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gProcessedSigError	rProcessedSigError	float		The universal or propagated error left after all the processing steps of Feature Extraction have been completed. In the case of one color, ProcessedSignalError has had the Error Model applied and will contain at least the larger of the universal (UEM) error or the propagated error. If multiplicative detrending is performed, ProcessedSignalError contains the error propagated from detrending. This is done by dividing the error by the normalized MultDetrendSignal.
gMedianSignal	rMedianSignal	float		Raw median signal of feature in green (red) channel (inlier pixels)
gBGMeanSignal	rBGMeanSignal	float		Mean local background signal (local to corresponding feature) computed per channel (inlier pixels)
gBGMedianSignal	rBGMedianSignal	float		Median local background signal (local to corresponding feature) computed per channel (inlier pixels)
gBGPixSDev	rBGPixSDev	float		Standard deviation of all inlier pixels per local BG of each feature, computed independently in each channel
gNumSatPix	rNumSatPix	integer		Total number of saturated pixels per feature, computed per channel
gIsSaturated	rIsSaturated	boolean	1 = Saturated or 0 = Not saturated	Boolean flag indicating if a feature is saturated or not. A feature is saturated IF 50% of the pixels in a feature are above the saturation threshold.
gIsLowPMTScaled Up	rIsLowPMTScaled Up	boolean	1 = Low 0 = High	Reports if the feature signal value is from the scaled-up low signal image or from the high signal image

Table 24 Feature results contained in the COMPACT output text file (COMPACT FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gIsFeatNonUnifOL	rIsFeatNonUnifOL	boolean	g(r)IsFeatNonUnifOL = 1 indicates Feature is a non-uniformity outlier in g(r)	Boolean flag indicating if a feature is a NonUniformity Outlier or not. A feature is non-uniform if the pixel noise of feature exceeds a threshold established for a "uniform" feature.
gIsBGNonUnifOL	rIsBGNonUnifOL	boolean	g(r)IsBGNonUnifOL = 1 indicates Local background is a non-uniformity outlier in g(r)	The same concept as above but for background.
gIsFeatPopnOL	rIsFeatPopnOL	boolean	g(r)IsFeatPopnOL = 1 indicates Feature is a population outlier in g(r)	Boolean flag indicating if a feature is a Population Outlier or not. Probes with replicate features on a microarray are examined using population statistics. A feature is a population outlier if its signal is less than a lower threshold or exceeds an upper threshold determined using a multiplier (1.42) times the interquartile range (i.e., IQR) of the population.
gIsBGPopnOL	rIsBGPopnOL	boolean	g(r)IsBGPopnOL = 1 indicates local background is a population outlier in g(r)	The same concept as above but for background
IsManualFlag		boolean		Lets you know if features have been flagged for downstream filtering in third party gene expression software.
gBGSubSignal	rBGSubSignal	float	g(r)BGSubSignal = g(r)MeanSignal - g(r)BGUsed	Background-subtracted signal. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.

Table 24 Feature results contained in the COMPACT output text file (COMPACT FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gIsPosAndSignif	rIsPosAndSignif	boolean	g(r)isPosAndSignif = 1 indicates Feature is positive and significant above background	Boolean flag, established via a 2-sided t-test, indicates if the mean signal of a feature is greater than the corresponding background (selected by user) and if this difference is significant. To view variables used in the t-test, see Table 33 on page 213.
gIsWellAboveBG	rIsWellAboveBG	boolean		Boolean flag indicating if a feature is WellAbove Background or not, feature passes g(r)IsPosAndSignif and additionally the g(r)BGSubSignal is greater than 2.6*g(r)BG_SD. You can change the multiplier 2.6.
gBGSDUsed	rBGSDUsed	float		Standard deviation of background used in g(r) channel; variable also used in t-test and surrogate algorithms. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, seeTable 33 on page 213.
SpotExtentX		float		Diameter of the spot (X-axis)
gTotalProbeSignal		float		This signal is the robust average of all the processed green signals for each replicated probe multiplied by the total number of probe replicates, the EffectiveFeature SizeFraction, the Nominal Spot Area and the Weight. For miRNA analyses
gTotalProbeError		float		This error is the robust average of all the processed green signal errors for each replicated probe multiplied by the total number of probe replicates, the EffectiveFeature SizeFraction, the Nominal Spot Area and the Weight. For miRNA analyses

Table 24 Feature results contained in the COMPACT output text file (COMPACT FEATURES table)*

Features (Green)	Features (Red)	Types	Options	Description
gTotalGeneSignal		float		This signal is the total probe signal times the number of probes per gene. For miRNA analyses
gTotalGeneError		float		This error is the square root of the square of the (total probe error times the number of probes per gene). For miRNA analyses
gIsGeneDetected		boolean		Lets you know if the gene was detected on the miRNA microarray.

* Results are reported to 9 decimal places in exponential notation for all result files.

Other text result file annotations

The following public accession numbers may or may not show up in the Feature Results section of the output text file.

Table 25 Public accession numbers in the output text file

Abbreviation	Description
dbj	DNA Database of Japan
emb	EMBL
gb	GenBank
gbpri	GenBank primate nucleotide accession number
gi	GenBank Gene Identifier
gp	GenPept protein identification number
mgi	Mouse Genome Informatics
pdb	Brookhaven Protein data bank
pir	NBRF PIR
prf	Protein Research Foundation
rafl	RIKEN full Length cDNA
ref	RefSeq
sp	SwissProt
tair	The Arabidopsis Information Resource
ug	UniGenelocuslink: LocusLink ID
wi	Whitehead

4
MAGE-ML (XML) File Results

This chapter provides a listing of MAGE-ML results in the form of tables. Refer to these tables when you want to know the results reported in a particular file. This chapter also contains a section on TIFF files and formats.

How Agilent output file formats are used by databases

Pattern files should be loaded to the database via FTP if possible to ensure that the pattern element, name attribute, is used to name the pattern.

Data analysis programs must match up information about the layout and annotation of the microarray features with the profile result files for each microarray within their databases. Agilent provides this design information for its microarrays in a variety of file formats, including GAL and MAGE-ML. These files describe the gene probes and their number and spacing on the microarray. Profile result files contain the signal and error information for each of the hybridized gene probes on the microarray.

Both pattern files and profile result files contain information that can be formatted in several ways: tab-delimited text format or an XML format, MAGE-ML.

Agilent only supports GEML2 Pattern files and MAGE-ML profiles for use with Rosetta Resolver. The pattern name in Rosetta Resolver should match the profile pattern name embedded in the profile data so that the data can be correctly associated. To do this, use the pattern autoimport function in Rosetta Resolver or correctly specify the pattern name when manually importing the pattern. (The Agilent pattern name in most cases is “Agilent-xxxxxx” where the xxxxxx is the AMADID number of the microarray.)

For transfer of data into GeneSpring, the pattern information can be obtined from within the Feature Extraction profile tab text file or can be obtained by download from the GeneSpring web site.

Tables of MAGE-ML results

Differences between MAGE-ML and text result files

The MAGE-ML result file includes most of the same parameters, statistics and results as the FULL text result file with the following differences:

- Scanner control parameters are included in the file.
- Some Feature Extraction parameter names (FE PARAMS table) have been changed to accommodate Rosetta Resolver terminology.
- MAGE result file includes all information included in the FEATURES table except for annotations, deletion control information and spot size information.
- Feature results (FEATURES table) are associated with quantitation types as defined by the Object Management Group in its Gene Expression Specification paper of February 2003 V.1. These types are listed below:
  - Measured Signal
  - Derived Signal
  - Ratio
  - Confidence Indicators—error and p-value
  - Specialized Quantitation Type (SQT) — includes all other data

Full and Compact Output Packages

In the Properties sheet for the project you can select if you want the MAGE-ML result file to contain all the possible columns and results (Full) or a reduced set of results (Compact).

MAGE-ML files can also be compressed before they are sent via FTP. Compressed MAGE-ML files further reduces the size of the file to decrease the transfer time. Use both Compact and

Compressed MAGE-ML files for Resolver. The Compact package contains only those columns required by Resolver, GeneSpring, CGH Analytics and Chip Analytics.

Tables for Full Output Package

Table 26 Scan protocol parameters in MAGE-ML result file

Parameter	Description
Image acquisition identifier	Barcode or identifier for microarray
Log information	Warnings and errors during run
Activity date	Time stamp for scanner run
Scanner information	Information such as name, make model and serial number of scanner
Operator	Person that runs scanner
ScanNumber	Number of the scan associated with the values listed in this table
Red.LASER_POWER_VALUE	Value of laser power in red channel
Green.LASER_POWER_VALUE	Value of laser power in green channel
Red.PMT_GAIN_VALUE	Photomultiplier gain in red channel
Green.PMT_GAIN_VALUE	Photomultiplier gain in green channel
Red.Saturation_Value	Signal value beyond which signal is saturated in the red channel
Green.Saturation_Value	Signal value beyond which signal is saturated in the green channel
MICRONS_PER_PIXEL_X	Radius of pixel in the x direction
MICRONS_PER_PIXEL_Y	Radius of pixel in the y direction
GlassThickness	Thickness of microarray slide
Red.DarkOffsetAverage	Dark offset data per image in red channel as measured by scanner

Table 26 Scan protocol parameters in MAGE-ML result file (continued)

Parameter	Description
Green.DarkOffsetAverage	Dark offset data per image in green channel as measured by scanner
PercentAutoFocusHold	Amount of movement in the autofocus because of fluctuations in the glass
DarkOffsetSubtracted	Resulting signal when dark offset value is subtracted

Table 27 Feature Extraction protocol parameters in MAGE-ML result file Differences between FEPARAMS in text file and MAGE-ML file

Text File FEPARAMS	MAGE-ML File FEPARAMS
Ratio_ErrorModel	Error Model
Ratio_AddErrorRed	Red.ADDITIVE_ERROR
Ratio_AddErrorGreen	Green.ADDITIVE_ERROR
Ratio_MultErrorRed	Red.MULTIPLICATIVE_ERROR
Ratio_MultErrorGreen	Green.MULTIPLICATIVE_ERROR

NOTE For 1-color, red signals and log ratios are not included in the MAGE-ML output files.

Table 28 Feature results (Full) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
SQT*	X_IMAGE_POSITION Y_IMAGE_POSITION			Found coordinates of the feature centroid
SQT	SpotExtentX SpotExtentY			Diameter of the spot (X- or Y-Axis)

Table 28 Feature results (Full) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
Ratio	LogRatio (base 10)			log(REDsignal/GREENsignal) per feature (processed signals used to calculate log ratio)
				If SURROGATES are turned off, then:
			-4	if DyeNormRedSig <= 0.0 & DyeNormGreenSig > 0.0
			4	if DyeNormRedSig > 0.0 & DyeNormGreenSig <= 0.0
			0	if DyeNormRedSig <= 0.0 & DyeNormGreenSig <= 0.0
Error	LogRatioError			If SURROGATES are turned off, then:
			1000	if DyeNormRedSig <= 0.0 OR DyeNormGreenSig <= 0.0
				IF SURROGATES are turned on, then:
				LogRatioError = error of the log ratio calculated according to the error model chosen
PValue	PValueLogRatio			Significance level of the Log Ratio computed for a feature
SQT	gSurrogateUsed	rSurrogateUsed	Non-zero value	The g(r) surrogate value used
			0	No surrogate value used

Table 28 Feature results (Full) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
SQT	gIsFound	rIsFound	1 = IsFound 0 = IsNotFound	A boolean used to flag found (strong) features. The flag is applied independently in each channel. A feature is considered found if the calculated spot centroid is within the bounds of the spot deviation limit with respect to corresponding nominal centroid. NOTE: IsFound was previously termed IsStrong.
Derived Signal	Green.DerivedSignal	Red.DerivedSignal		The propagated feature signal, per channel, used for computation of log ratio
Error	Green.ProcessedSig Error	Red.ProcessedSig Error		Standard error of propagated feature signal, per channel
SQT	gNumPixOLHi	rNumPixOLHi		Number of outlier pixels per feature with intensity > upper threshold set via the pixel outlier rejection method. The number is computed independently in each channel. These pixels are omitted from all subsequent calculations.
SQT	gNumPixOLLo	rNumPixOLLo		Number of outlier pixels per feature with intensity < lower threshold set via the pixel outlier rejection method. The number is computed independently in each channel. NOTE: The pixel outlier method is the ONLY step that removes data in Feature Extraction.
SQT	gNumPix	rNumPix		Total number of pixels used to compute feature statistics, i.e., total number of inlier pixels/per spot, same in both channels

Table 28 Feature results (Full) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
Measured Signal	Green.Measured Signal	Red.Measured Signal		Raw mean signal of feature in green (red) channel
SQT	gMedianSignal	rMedianSignal		Raw median signal of feature in green (red) channel
SQT	gNetSignal	rNetSignal		MeanSignal minus DarkOffset
Error	Green.PixSDev	Red.PixSDev		Standard deviation of all inlier pixels per feature. This is computed independently in each channel.
SQT	gBGNumPix	rBGNumPix		Total Number of pixels used to compute Local BG statistics per spot; i.e., total number of BG inlier pixels. This number is computed independently in each channel.
Measured Signal	Green.Background	Red.Background		Mean local background signal (local to corresponding feature) computed per channel
SQT	gBGMedianSignal	rBGMedianSignal		Median local background signal (local to corresponding feature) computed per channel
Error	Green.BGPixSDev	Red.BGPixSDev		Standard deviation of all inlier pixels per Local BG of each feature, computed independently in each channel
SQT	gNumSatPix	rNumSatPix		Total number of saturated pixels per feature, computed per channel
SQT	gIsSaturated	rIsSaturated	1 = Saturated or 0 = Not saturated	Integer indicating if a feature is saturated or not. A feature is saturated IF 50% of the pixels in a feature are above the saturation threshold.
SQT	gIsLowPMTScaledUp	rIsLowPMTScaledUp	1 = Low 0 = High	For XDR features, this is an integer indicating if the low PMT value was used for the calculations, or the high value.

Table 28 Feature results (Full) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
SQT	PixCorrelation			Ratio of estimated feature covariance in RedGreen space to product of feature Standard Deviation in Red Green space The covariance of two features measures their tendency to vary together, i.e., to co-vary. In this case, it is a cumulative quantitation of the tendency of pixels belonging to a particular feature in Red and Green spaces to co-vary.
float	BGPixCorrelation			The same concept as above but in case of background
SQT	gIsFeatNonUnifOL	rIsFeatNonUnifOL	g(r)IsFeatNonUnifOL = 1 indicates Feature is a non-uniformity outlier in g(r)	Integer indicating if a feature is a NonUniformity Outlier or not. A feature is non-uniform if the pixel noise of feature exceeds a threshold established for a "uniform" feature.
SQT	gIsBGNonUnifOL	rIsBGNonUnifOL	g(r)IsBGNonUnifOL = 1 indicates Local background is a non-uniformity outlier in g(r)	The same concept as above but for background
SQT	gIsFeatPopnOL	rIsFeatPopnOL	g(r)IsFeatPopnOL = 1 indicates Feature is a population outlier in g(r)	Boolean flag indicating if a feature is a Population Outlier or not. Probes with replicate features on a microarray are examined using population statistics. A feature is a population outlier if its signal is less than a lower threshold or exceeds an upper threshold determined using a multiplier (1.42) times the interquartile range (i.e., IQR) of the population.

Table 28 Feature results (Full) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
SQT	gIsBGPopnOL	rIsBGPopnOL	g(r)IsBGPopnOL = 1 indicates local background is a population outlier in g(r)	The same concept as above but for background
SQT	IsManualFlag			
SQT	gBGSubSignal	rBGSubSignal	gBGSubSignal = gMeanSignal - gBGUsed	Background-subtracted signal To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.
Error	gBGSubSigError	rBGSubSigError		Propagated standard error as computed on net g(r) background-subtracted signal
SQT	BGSubSigCorrelation			Ratio of estimated background-subtracted feature signal covariance in RG space to product of background-subtracted feature Standard Deviation in RG space
SQT	gIsPosAndSignif	rIsPosAndSignif	g(r)isPosAndSignif = 1 indicates Feature is positive and significant above background	Boolean flag, established via a 2-sided t-test, indicates if the mean signal of a feature is greater than the corresponding background (selected by user) and if this difference is significant. To view variables used in the t-test, see Table 33 on page 213.
SQT	gPValFeatEqBG	rPValFeatEqBG		P-value from t-test of significance between g(r)Mean signal and g(r) background
SQT	gIsWellAboveBG	rIsWellAboveBG		Boolean flag indicating if a feature is WellAbove Background or not Feature passes g(r)IsPosAndSignif and additionally the g(r)BGSubSignal is greater than 2.6*g(r)BGSDUsed.

Table 28 Feature results (Full) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
Boolean	gSpatialDetrendIsIn FilteredSet	rSpatialDetrendIsIn FilteredSet		Set to true for a given feature if it is part of the filtered set used to detrend the background. This feature is considered part of the locally weighted lowest x% of features as defined by the *DetrendLowPassPercentage*.
float	gSpatialDetrend SurfaceValue	rSpatialDetrend SurfaceValue		Value of the smoothed surface calculated by the Spatial detrend algorithm
SQT	IsUsedBGAdjust		1 = Feature used 0 = Feature not used	A boolean used to flag features used for computation of global BG offset
SQT	gBGUsed	rBGUsed	gBGSubSignal = gMeanSignal - gBGUsed	Background used to subtract from the MeanSignal; variable also used in t-test. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.
Error	gBGSDUsed	rBGSDUsed		Standard deviation of background used in g(r) channel; variable also used in t-test and surrogate algorithms. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.
SQT	IsNormalization		1 = Feature used; 0 = Feature not used	A boolean flag that indicates if a feature is used to measure dye bias
SQT	Green.DyeNorm Signal	Red.DyeNormSignal		The dye-normalized signal in the indicated channel
Error	Green.DyeNormError	Red.DyeNormError		The standard error associated with the dye-normalized signal
SQT	DyeNormCorrelation			Dye-normalized red and green pixel correlation

Table 28 Feature results (Full) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
SQT	ErrorModel		0 = Propagated model chosen by you or by software 1 = Universal error model chosen by you or by software	Indicates the error model that you chose for Feature Extraction or that the software uses if you have chosen the "Most Conservative" option
SQT	xDev			A signal-to-noise parameter used to calculate p-value; calculated differently depending on error model chosen
Failed	Failed			Attached to any feature result that does not meet the criteria described in "XML Control Type output" on page 181

* SQT — Specialized Quantitation Type

Table for Compact Output Package

This table contains only those columns required by Resolver, GeneSpring, CGH Analytics and Chip Analytics.

In the Compact version of the MAGE-ML file, the entire FEPARAMS section is included. MAGE-ML has a rich mechanism for describing protocols and protocol parameters.

Table 29 Feature results (Compact) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
Ratio	LogRatio (base 10)			log(REDsignal/GREENsignal) per feature (processed signals used to calculate log ratio)
				If SURROGATES are turned off, then:
			-4	if DyeNormRedSig <= 0.0 & DyeNormGreenSig > 0.0
			4	if DyeNormRedSig > 0.0 & DyeNormGreenSig <= 0.0
			0	if DyeNormRedSig <= 0.0 & DyeNormGreenSig <= 0.0
SQT*	X_IMAGE_POSITION Y_IMAGE_POSITION		float	Found coordinates of the feature centroid in microns

Table 29 Feature results (Compact) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
Error	LogRatioError			If SURROGATES are turned off, then: if DyeNormRedSig <= 0.0 OR DyeNormGreenSig <= 0.0 IF SURROGATES are turned on, then: LogRatioError = error of the log ratio calculated according to the error model chosen
			1000	
PValue	PValueLogRatio			Significance level of the Log Ratio computed for a feature
Derived Signal	Green.DerivedSignal	Red.DerivedSignal		The propagated feature signal, per channel, used for computation of log ratio
Error	Green.ProcessedSig Error	Red.ProcessedSig Error		Standard error of propagated feature signal, per channel
Measured Signal	Green.Measured Signal	Red.Measured Signal		Raw mean signal of feature in green (red) channel
SQT	gMedianSignal	rMedianSignal		Raw median signal of feature in green (red) channel
SQT	gBGMedianSignal	rBGMedianSignal		Median local background signal (local to corresponding feature) computed per channel
Error	Green.BGPixSDev	Red.BGPixSDev		Standard deviation of all inlier pixels per Local BG of each feature, computed independently in each channel
SQT	gIsSaturated	rIsSaturated	1 = Saturated or 0 = Not saturated	Integer indicating if a feature is saturated or not. A feature is saturated IF 50% of the pixels in a feature are above the saturation threshold.

Table 29 Feature results (Compact) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
SQT	gIsLowPMTScaledUp	rIsLowPMTScaledUp	1 = Low 0 = High	For XDR features, this is an integer indicating if the low PMT value was used for the calculations, or the high value.
SQT	gIsFeatNonUnifOL	rIsFeatNonUnifOL	g(r)IsFeatNonUnifOL = 1 indicates Feature is a non-uniformity outlier in g(r)	Integer indicating if a feature is a NonUniformity Outlier or not. A feature is non-uniform if the pixel noise of feature exceeds a threshold established for a "uniform" feature.
SQT	gIsBGNonUnifOL	rIsBGNonUnifOL	g(r)IsBGNonUnifOL = 1 indicates Local background is a non-uniformity outlier in g(r)	The same concept as above but for background
SQT	gIsFeatPopnOL	rIsFeatPopnOL	g(r)IsFeatPopnOL = 1 indicates Feature is a population outlier in g(r)	Boolean flag indicating if a feature is a Population Outlier or not. Probes with replicate features on a microarray are examined using population statistics. A feature is a population outlier if its signal is less than a lower threshold or exceeds an upper threshold determined using a multiplier (1.42) times the interquartile range (i.e., IQR) of the population.
SQT	gIsBGPopnOL	rIsBGPopnOL	g(r)IsBGPopnOL = 1 indicates local background is a population outlier in g(r)	The same concept as above but for background
SQT	gBGSubSignal	rBGSubSignal	gBGSubSignal = gMeanSignal - gBGUsed	Background-subtracted signal To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.

Table 29 Feature results (Compact) contained in the MAGE-ML (FEATURES table)

Quant Type	Features (Green)	Features (Red)	Options	Description
SQT	IsManualFlag			Boolean flag that describes if the feature centroid was manually adjusted.
SQT	gIsPosAndSignif	rIsPosAndSignif	g(r)isPosAndSignif = 1 indicates Feature is positive and significant above background	Boolean flag, established via a 2-sided t-test, indicates if the mean signal of a feature is greater than the corresponding background (selected by user) and if this difference is significant. To view variables used in the t-test, see Table 33 on page 213.
SQT	gIsWellAboveBG	rIsWellAboveBG		Boolean flag indicating if a feature is WellAbove Background or not Feature passes g(r)IsPosAndSignif and additionally the g(r)BGSubSignal is greater than 2.6*g(r)BGSDUsed.

* SQT — Specialized Quantitation Type

Helpful hints for transferring Agilent output files

XML output

There are several situations you should be aware of as you use MAGE-ML (XML) output with gene expression data analysis software from Rosetta BioSoftware (Rosetta Resolver® or Rosetta Luminator™ software):

If there is no barcode

If there is no barcode in the original .tif file for whatever reason, there will be no barcode information in the MAGE-ML output (warning message in Project Run summary). For the data to load into Rosetta Resolver, it must have a barcode associated with it. You can add barcode information in the Scan Image Properties dialog box. See "Display file information" on page 215 of the *User Guide*.

Access control list (ACL)

Rosetta Resolver knows about the access control list (ACL) assigned to the scan and can easily recognize and load any MAGE-ML file. The owner of the data sets the chip and hybe access controls in Rosetta Resolver before importing the profile (scan) data. For autoimport, the profile is normally placed in the **MAGE** directory.

XML Control Type output

If a feature is used in dye normalization, its Control_Type is normalization, even though it can also be a positive or negative control. If a feature is not used in normalization, it is either positive, negative, deletion, mismatch, or false.

Table 30 Control Type Definitions

Name	XML
Probe	false
Positive	pos or positive
Negative	neg or negative
Not Probe*	notprobe

*Not Probe—These features are feature extracted, but they are not used by Feature Extraction as input to any calculations; these features are not used during outlier analysis or for the dye normalization calculation. However, dye normalization values and ratios are calculated, and the results appear in the text and XML output files, and the feature extraction visual results file. An exception is that Not Probe's background is used in the calculation of the local background with the radius method.

Conversion of feature flag information

Failed (MAGE-ML) produce the following settings:

- Bit 8 (green) and 12 (red) are set if the feature is saturated in both channels.
- Bit 18 is set if the feature, or its deletion control, is a non-uniformity outlier in either color.
- Bit 23 is set if the probe is low specificity, e.g., when the deletion control is greater than or equal to the feature.

TIFF Results

You can transfer the original TIFF file or a JPEG file to Rosetta Resolver or a third-party program. The shape file, .shp, created during Feature Extraction cannot be viewed by any program other than Agilent Feature Extraction software.

TIFF file format options

See "Display file information" on page 215 of the User Guide for more information on the File Info dialog box.

Feature Extraction supports the TIFF file format. All file information for each file is listed in the File Info dialog box. The TIFF file is compliant with Adobe version 6.0 file format. The complete specification is available from the following URL: http://partners.adobe.com/asn/developer/PDFS/TN/TIFF6.pdf.

There are two sets of custom TIFF tags in the Agilent file format.

Genetic Analysis Technology Consortium (GATC) TIFF Tags

Agilent Technologies is not a member of GATC or otherwise connected to this organization, and makes no internal use of these tags. They are included for the convenience of customers who use software that requires them.

Custom TIFF Tags Agilent Technologies uses its own custom TIFF tags for storing additional file information.

TIFF Tag 37701 This tag points to a data structure. This data structure is not public, but information stored in the data structure is available to customers in the MATLAB file format.

TIFF Tag 37702 This tag points to a string containing the file description. The usual TIFF description tags (tag 270) are used to hold the color name, "red" or "green," for each image. This allows programs that interpret only "standard" TIFF tags to determine image colors. The Page Name tag (tag 285) also contains the color names.

5
How Algorithms Calculate Results

This chapter shows you how each Feature Extraction algorithm uses its parameters to calculate results that are passed on to the next algorithm and finally on to third-party data analysis programs.

Overview of Feature Extraction algorithms

Protocol step algorithms operate similarly during the Feature Extraction process for 2-color gene expression, CGH and non-Agilent microarrays. That is, the algorithms and parameter fields are similar, but the parameter values are different depending on the protocol. The Feature Extraction process for 1-color gene expression microarrays includes only six protocol steps, and for miRNA analysis the process includes those six steps plus a MicroRNA Analysis step.

The algorithms described below and on the following pages are primarily for the Feature Extraction process of 2-color microarrays. Any differences in algorithms and functions for other microarray experiments are also explained.

Algorithms and functions they perform

Place Grid

This algorithm finds the grid to define the nominal positions of the spots on the microarray.

For more information on the algorithms for XDR extraction, see "XDR Extraction Process" on page 196.

eXtended Dynamic Range (XDR) extraction For an XDR extraction, the grid placement is done using the high intensity scan (i.e., higher PMT voltage). The grid found using the high intensity scan is used as the starting point for the remaining extraction of both the high and low intensity images.

Find Spots

This algorithm locates the exact size and centroid of each spot on the scanned microarray. Once the spot centroids have been located, the CookieCutter algorithm or WholeSpot algorithm defines the feature for each spot. The software then defines the local background for each spot based on the radius of a circle drawn around the spot.

Next, the pixel outlier algorithm identifies outlier pixels in the feature and in the local background for each spot. These pixels are then omitted from further calculations. This is the only point where data is omitted. Subsequent outlier analyses flag data, but do not remove the data.

Inlier pixels within the cookie area represent a feature while the inlier pixels within the annulus around the feature, after excluding the exclusion zone, represent the local background. The Feature Extraction program calculates the following values from these inlier pixels: mean, median, standard deviation, normalized IQR, and number of inlier pixels.

XDR extraction This is the only step that is run twice on an XDR extraction. The spot placement and spot measurements are found separately for the high and low intensity scans. Then the XDR algorithm decides on a feature by feature basis which scan the data should come from (more on this below). For features that are very bright in the high intensity scan, the XDR algorithm uses the data from the low intensity scan. This choice is made independently for each color channel.

For each feature that uses data from the low intensity scan, the following columns get replaced (determined separately for red and green channels): NumPixOLHi, NumPixOLLo, NumPix, MeanSignal, MedianSignal, PixSDev, PixNormIQR, NumSatPix, IsSaturated, NetSignal.

These columns include the raw data from the spotfinding and measurement steps (signal levels, pixel noise levels, number of pixels, if the pixels and feature are saturated). Once the substitutions have been made to some features in each color channel, the extraction proceeds as if there were only a single combined set of features.

Flag Outliers

Next, the Flag Outliers algorithm flags anomalous features and local backgrounds as non-uniformity outliers and/or population outliers. Population outlier flagging is based on population statistics of replicate features on the microarray.

Which of two statistical tests is used to identify population outliers depends on the number of replicaate features on the microarray.

Non-uniformity outlier flagging is based on statistical deviation from the expected noise in the Agilent microarray-based Gene Expression system (scanner, labeling/hybridization protocols, and microarrays). The algorithm automatically calculates the B (linear) and C (constant) terms of the polynomial fit for the expected noise for any type of microarray experiment.

Compute Bkgd, Bias and Error

This algorithm applies background subtraction to each feature to yield the background-subtracted intensity. You can also apply a “spatial detrend” algorithm to estimate and remove noise due to a systematic gradient on the microarray.

Another algorithm can correct for any underestimation or overestimation of the background in both the red and green channels of low-intensity signals by applying a global background adjustment value to the background-subtracted signals.

Before using the algorithm for estimating the error, the system uses an algorithm to calculate robust negative control statistics for both CGH and miRNA data. CGH microarrays have a variety of sequences that are used as negative controls. Occasionally, “hot” features are not flagged as population outliers. In addition, “hot” sequences may exist; that is, all features of that sequence have higher signals than features in other negative control sequences. These problems can inflate NegC SD, which is used in the calculation of AdditiveError for the CGH error model.

To provide an estimate of the error in the background-subtracted signal calculation, the error model is now calculated after background subtraction. The 1-color error model has been changed to exactly mimic the 2-color error model.

To determine if the feature intensity is significant compared to the background intensity, two kinds of tests are available: t-test and WellAboveBG test. Both of these tests depend upon an estimation of background error.

The default protocol for older Agilent protocols still uses pixel statistics of local background regions to estimate background error in the 2-sided t-test. Newer Agilent protocols use an improved estimation of background error: the additive error, calculated from the Agilent error model. You can choose between these two background error estimations in the protocol parameter field, “Significance (for IsPosAndSignif and IsWellAboveBG)”.

The WellAboveSDMulti confidence test is used to determine if the feature background-subtracted signal is well above its background error.

Surrogates are calculated here and depend on the significance model used. Given the standard t-test the surrogates are calculated exactly as before. Given the new significance test based upon additive error, the surrogate value is determined by the additive error and the p-value.

The program can also use a multiplicative detrend algorithm, if selected or the default in the protocol, to provide a surface fit to account for the dome effect that can happen when microarrays are processed. Multiplicative detrending is turned on by default for all v.9.5 protocols, except for the non-Agilent protocol.

Placing the error model calculation step before the significance calculation permits the result of the error model calculation to be used for the significance calculation, surrogate calculation and multiplicative detrending steps.

Correct Dye Biases

Since dye bias between the red and green channels is a common phenomenon in a dual-color microarray platform, this algorithm adjusts for the bias by multiplying the background-subtracted signals with the appropriate dye normalization factors. Both linear and non-linear (locally weighted) normalization methods are available.

Surrogates are applied after the dye norm fit and before the dye normalization takes place. This ensures that only real data contribute to the fit and also surrogate data is correctly dye-normalized for both the Linear and Lowess options.

Because 1-color experiments use only the green channel, they do not use this protocol step. Surrogates exist and can be used for 1-color. Since there is no dye norm for 1-color, surrogates are applied at the end of Compute Background Bias and Error.

Compute Ratios

This algorithm determines if a feature is differentially expressed by calculating the log ratio of the red over green processed signals. The processed signal is the dye-normalized signal, except in situations when the raw mean signal is less than the background or not significant compared to the background, or the background-subtracted signal is less than its background standard deviation.

Because 1-color experiments use only the green channel, they do not use this protocol step.

Calculate Metrics

These algorithms calculate all the QC metrics for the analysis. One of the primary algorithms in this step is the gridding test, whose parameter values are hidden in the protocol. This algorithm yields grid warnings on the Summary Reports and the “Evaluate Grid” warning in the QC Report. In v.9.5, Agilent has added many more tests to assess if gridding has been successful or not.

MicroRNA Analysis

This step is used in the 1-color miRNA analysis after background effects have been accounted for. The algorithms in this step calculate the TotalGeneSignal and the TotalGeneError for the analysis.

Generate Results

This part of the process generates the output result files using the parameter values specified in the protocol step and the selections made in the Project Properties window. This step is not discussed in this chapter.

Algorithms and results they produce

The table below summarizes the results for each algorithm (protocol step). These result names are used in the equations for the calculations for each algorithm.

Table 31 Algorithms (Protocol Steps) and the results they produce

Protocol Step	Results	Result Definition
Find Spots	MeanSignal	Average raw signal of feature calculated from the intensities of all inlier pixels that represent the feature (after outlier pixel rejection). The number of inlier pixels is shown in the column NumPix.
Find Spots	MedianSignal	Median raw signal of feature calculated from the intensities of all inlier pixels that represent the feature (after outlier pixel rejection). The number of inlier pixels is shown in the column NumPix.
Find Spots	BGMeanSignal	Average raw signal of the local background calculated from intensities of all inlier pixels that represent the local background of the feature (after outlier pixel rejection). The number of inlier pixels is shown in the column BGNumPix.
Find Spots	BGMedianSignal	Median raw signal of the local background calculated from intensities of all inlier pixels that represent the local background of the feature (after outlier pixel rejection). The number of inlier pixels is shown in the column BGNumPix.
Find Spots	NetSignal	MeanSignal minus Dark Offset
Find Spots	IsSaturated	A Boolean flag of 1 indicates that the feature is saturated; at least 50% of the inlier pixels in the feature have intensities above the saturation threshold. One can determine the saturation level of a feature by dividing the NumSatPix by the NumPix.
Flag Outliers	IsFeatureNonUnifOL	A Boolean flag of 1 indicates that the feature is a non-uniformity outlier; the measured feature pixel variance is greater than the expected feature pixel variance plus the confidence interval.
Flag Outliers	IsFeatPopOL	A Boolean flag of 1 indicates that the feature is a population outlier. This means that the feature MeanSignal is greater than the upper rejection boundary or less than the lower rejection boundary, both of which are determined by multiplying a factor (1.42) by the interquartile range of the population, made up of intra-array feature replicates. (See "Step 5: Reject outliers" on page 204.)

Table 31 Algorithms (Protocol Steps) and the results they produce

Protocol Step	Results	Result Definition
Compute Bkgd, Bias and Error	BGAdjust	An adjustment value added to the initial background-subtracted signal to correct for underestimation or overestimation of the background. This value can be positive or negative. Note the BGAdjust values are reported per channel in the STATS table of Feature Extraction text file.
Compute Bkgd, Bias and Error	BGused	Final background signal used to subtract the background from the feature mean signal. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.
Compute Bkgd, Bias and Error	BGSubSignal	Feature signal after subtraction of the background corrections. To view the values used to calculate this variable using different background signals and settings of spatial detrend and global background adjust, see Table 33 on page 213.
Compute Bkgd, Bias and Error	IsPosAndSignif	If significance is based on pixel statistics, a Boolean flag of 1 indicates that the feature MeanSignal is greater than and significant compared to the background signal (i.e BGUsed). If significance is based on the Additive Error of the Error Model, a Boolean flag of 1 means that the feature MeanSignal is greater than and significant compared to the Additive Error,
Compute Bkgd, Bias and Error	IsWellAboveBG	A Boolean flag of 1 indicates that the feature BGSubSignal is well above background and passes the IsPosAndSignif test.
Compute Bkgd, Bias and Error	SpatialDetrendIsIn FilteredSet	Set to true for a given feature if it is part of the filtered set used to detrend the background. The feature may be in the set of locally weighted lowest x% of features as defined by the *DetrendLowPassPercentage,* may be a negative control feature or may be part of the set of features that are in the negative control range. The feature set is defined by the detrend method selected.
Compute Bkgd, Bias and Error	SpatialDetrend SurfaceValue	Value of the smoothed surface, at that feature, calculated by the Spatial detrend algorithm

Table 31 Algorithms (Protocol Steps) and the results they produce

Protocol Step	Results	Result Definition
Compute Bkgd, Bias and Error	MultDetrend Signal	A surface is fitted through the log of the background-subtracted signal to look for multiplicative gradients. A normalized version of that surface interpolated at each point of the microarray is stored in MultDetrendSignal. The surface is normalized by dividing each point by the overall average of the surface. That average is stored in MultDetrendSurfaceAverage as a statistic. If the protocol uses the option to fit to only replicate features, the surface is normalized for the fit. The MultDetrend SurfaceAverage is smaller in this case, a number around 1.
Compute Bkgd, Bias and Error	SurrogateUsed	A non-zero surrogate value indicates that the MeanSignal is less than or not significant versus the background or the BGSubSignal is less than its background standard deviation. To determine how the background SD is calculated, see Table 33 on page 213.
Correct Dye Biases	DyeNormSignal	A dye-normalized signal calculated by multiplying the BGSubSignal with the appropriate DyeNormFactor.
Correct Dye Biases	LinearDyeNormFactor (Table 20 on page 113)	A global constant to normalize the dye bias from all feature background-subtracted signals. LinearDyeNormFactor is calculated such that geometric mean intensity of the selected normalization features equals 1000.
Compute Ratios	ProcessedSignal	The signal left after all the FE processing steps have been completed. In the case of 1-color, ProcessedSignal contains the Multiplicatively Detrended BackgroundSubtracted Signal if the detrending is selected and helps. If the detrending does not help, this column will contain the BackgroundSubtractedSignal.
Compute Ratios	ProcessedSigError	The universal or propagated error left after all the processing steps of the Feature Extraction process have been completed. In the case of one color, If multiplicative detrending is performed, ProcessedSignalError contains the error propagated from detrending. This is done by dividing the error by the normalized MultDetrendSignal.
Compute Ratios	LogRatio	Log of the ratio of rProcessedSignal over gProcessedSignal. The log ratio indicates the level of gene expression in cyanine 5-labeled sample relative to cyanine 3-labeled sample.

Table 31 Algorithms (Protocol Steps) and the results they produce

Protocol Step	Results	Result Definition
Compute Ratios	pValueLogRatio	P-value indicates the level of significance in the differential expression of a gene as measured through the log ratio.
MicroRNA Analysis	gTotalGeneSignal	This signal is the total probe signal times the number of probes per gene. For miRNA analyses
MicroRNA Analysis	gTotalGeneError	This error is the square root of the square of the (total probe error times the number of probes per gene). For miRNA analyses

XDR Extraction Process

What is XDR scanning?

The Agilent scanner can cover a dynamic intensity range greatly in excess of the range covered by a single scan. Furthermore, Agilent microarray features can produce signals that span a broader range of intensity than a single scan can cover. Therefore, you can use eXtended Dynamic Range (XDR) to cover the full dynamic intensity range of your microarray features and hence see the most useful biology.

To do this you set the scanner to scan twice, once at a high PMT setting (the high intensity scan) followed immediately by a low PMT setting (the low intensity scan). This functionality is enabled using Agilent Scan Control Software version 7.0. The two scans are labeled in their tiff headers as paired scans of the same microarray.

XDR Feature Extraction process

The Feature Extraction program (9.1) uses this information to know to extract them as a pair. In this XDR extraction type, the Feature Extraction program processes the two scans together and produces a single set of outputs that contain data from both scans.

Some of the features contain data from the high intensity scan and some from the low intensity scan. You can determine this by viewing the column, r,gIsLowPMTScaledUp, for each color channel. For signals that are very bright (or saturated) in the high intensity scan (e.g., a scan at 100% PMT gain), the XDR algorithm substitutes the data from the low intensity scan (e.g., 10% PMT gain) after scaling the intensity appropriately.

To extract these arrays the Feature Extraction program uses a somewhat different flow of the image processing and data analysis algorithms.

The Feature Extraction program places the grid on the high intensity scan only, then finds spots using this grid on each of the two scans.

The XDR algorithm decides which features should use the low intensity scan data, scales these signals appropriately and does a replacement for each feature and color channel where appropriate. Then FE proceeds with the rest of the data analysis (outlier detection, background correction, dye normalization, etc.) exactly as it would for a single non-XDR scan.

Upon completion, the Feature Extraction program generates results as if they were from a single measurement of the microarray. The QC report and the stats table indicate that the Feature Extraction program extracted an XDR image pair by stating the new saturation value. This is the saturation value of the low intensity scan after suitable scaling. For instance, if the high intensity scan is at 100% and the low intensity scan is at 10%, the new saturation values will be around 650,000 (about 10x greater than a normal 100% PMT gain scan). This lets you use data in your calculations covering a much greater dynamic range.

How the XDR algorithm works

How does the XDR algorithm decide how to combine and scale the data from the high intensity and low intensity scans? The general theory is that the high intensity gives the best results for the low end of the signal range and the low intensity scan gives better data for bright features (less affected by saturation). The Feature Extraction program uses a signal level of 20,000 as the cut-off between the two scans. If the NetSignal of the high intensity scan is greater than 20,000 counts, then the data from the low intensity scan is used.

The low intensity scan is scanned with a lower PMT gain than the high intensity scan (say 10% versus 100%). So to combine the data the signals from the low intensity scan needs to increased to match those from the high intensity scans. To

determine the factory by which the low-intensity signal should be scaled, the algorithm uses features that have signals in an overlap range where both the high and low intensity scans provide very stable data. This range is Net Signals in the high intensity scan greater than 300 counts and less than 20,000 counts.

Using data in this range, the Feature Extraction program generates a linear fit (with a slope and an intercept) that transforms the low-intensity mean signals into the same range as high intensity scans. The final scaled signal for the XDR extraction is MeanSignal ([low-intensity scan * slope] + intercept).

The linear fit constants determined in this step are included in the stats table.

For signals over 20,000 counts in the high intensity scan, therefore, the low intensity scan signals can extend to nearly 1.2 million counts.

If the low intensity scan has a spot centroid too far from the high intensity centroid (greater than 2 pixels), the algorithm does not make a substitution.

Troubleshooting the XDR extraction

The XDR algorithm provides warnings in the project summary report to indicate an issue with the XDR extraction process.

- No XDR signal substitution for color red/green.

  This message appears if there are no features for which the low intensity data are substituted. This could occur on a dim array

- Computation of the XDR fit for red/green is based on only X pairs of (high PMT, low PMT) matching values.

  This message appears if very few features had data in the overlap range for the fit. The user should check the data in this case to confirm that the XDR combination is satisfactory.

- Computation of the XDR fit for red/green results in a large intercept.

  This message appears if the linear fit between the low and high intensity scans has a very large intercept.

  This can be indicative of a poor linear fit. The user should check the data in this case to confirm that the XDR combination is satisfactory.

- Computed XDR ratio for red/green is X vs expected Y from PMT settings. Check scanner calibration.

  This message appears if the ratio of the high/low intensity scans is different from what is expected from the scanner. For instance, an XDR scan set with 100% and 10% for PMT gain settings should yield a ratio close to 10. If this ratio is different than expected, the Feature Extraction program may or may not have performed correctly. But you should check the data in this case to confirm that the XDR combination is satisfactory.

  This message is more likely to appear as the low intensity PMT gain setting gets closer 1%. This is because the percentage error in the PMT gain setting increases as the setting moves away from 100%.

How each algorithm calculates a result

Place Grid

Step 1: Place a grid to find the nominal spot positions

After the Feature Extraction program automatically determines the format of the grid, it initiates the next steps.

The algorithm reduces the two-dimensional image data of the microarray to two one-dimensional data sets that are further processed to determine the layout of the grid on the microarray.

Projection of the two-dimensional microarray is performed to produce two one-dimensional datasets (projected signals). From the one-dimensional datasets peaks of the projected signals are filtered to determine which peaks to retain for further processing, based on predetermined peak height and peak width thresholds.

Nominal spacing between the features may be estimated based on a statistical determination of a most frequent distance between centers of retained peaks that are adjacent to one another. Coordinates for the features on the microarray, relative to the X and Y axes, are generated based on the selected peaks and peak spacing. The grid is then adjusted for rotation and skew.

Find Spots

Step 2: Locate the spot centroids

The calculation is based on an iterative Bayesian-probability-based pixel classification. A binary feature mask is created that classifies the pixels in a region of interest around each grid position into feature pixels or background pixels. The

approximate radius of each feature mask is considered as the corresponding spot radius and the center of mass of the feature mask is considered as the actual spot centroid.

In the visual results view (.shp file), all spots that are found are shown using a blue “X” on the spot and marked as “Found”. For all spots, the blue cross (+) shows the location of the grid. If the centroid cannot be found because the spot is too weak, or the distance between + and X centroids exceeds the range specified by the Spot Deviation Limit, this spot is labeled “Not Found”.

Step 3: Define features

See “Select a spot statistics method to define features” on page 159” of the *User Guide* for how the Feature Extraction program defines features either with the CookieCutter method or the WholeSpot method.

Step 4: Estimate the radius for the local background

The radius is the distance from the center of the cookie or whole spot to the edge of the outermost region, as shown in Figure 38. The default radius is the value specified in the protocol. You can also enter a minimum radius whose value is less than the default radius, or you can enter a larger radius to capture more pixels in the background. You can use the radius method for estimating global backgrounds as well.

The figures in this step represent the local background for the CookieCutter method for defining features. The radius for the local background is estimated in the same way for the WholeSpot method.

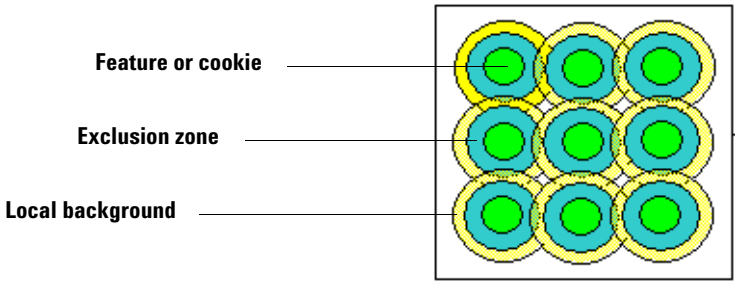


Figure 38 Local background in relation to other zones for CookieCutter method

Default radius The default radius is the radius of the local background for one feature. This radius is known as the SELF radius and its value is the default value that you see in the Find and Measure Spots protocol step if autoestimation is turned off.

Although the radius can map a circle that appears to overlap other features, the Feature Extraction program does not use these pixels to calculate the local background signal.

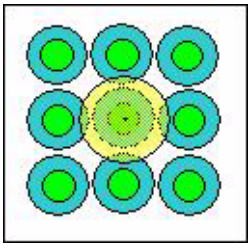


Figure 39 Example of a SELF radius

The value of the default radius (in microns) depends on the scan resolution and interspot spacing found in the TIFF and grid template or file, shown in equation [1]:

$$Default\ Local\ Radius = SELF = (0.6 \times Scan\_resolution \times Max\ (Interspotspacing\_x, Interspotspacing\_y)) \quad \textbf{[1]}$$

For the WholeSpot method, if extraction stops at this step, you may need to enter a larger radius than the protocol default radius.

The software autoestimates the Default Local Radius if specified in the protocol. Otherwise, you can enter this radius in the FE Protocol Editor.

Minimum radius The minimum radius that you can enter is the FLOOR (Default Radius), where FLOOR rounds the calculated value of the default radius down to the next lower integer, e.g., FLOOR (87.6) = 87.

Maximum radius The software lets you enter a maximum radius for the local background no greater than the distance from the center of the innermost feature to the edge of a circle that approximately surrounds the fourth closest set of nearest neighbors, or n=4, as shown in Equation 2. The set of eight nearest neighbors closest to the feature of interest is defined as n=1, as shown in Equation 3.

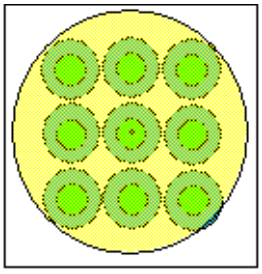


Figure 40 Example of the radius for the first closest set of nearest neighbors, or n=1 (eight nearest neighbors)

The value of the maximum radius also depends on the scan resolution and interspot spacing in the TIFF and grid template or file, shown in the equation below.

$$Max\ radius = \text{CEILING}\left[(Scan\_resolution \times 4.7)\ \sqrt{(Interspotspacing\_x)^2 + (Interspotspacing\_y)^2}\right] \quad \textbf{[2]}$$

where CEILING rounds the calculated value up to the next higher integer, e.g., CEILING [3.2] = 4.

Any radius The value of any radius between the minimum and maximum that circumscribes a circle surrounding the nth closest set of nearest neighbors from the central spot can be approximated as:

$$Radius\_n = Scan\_resolution \times n.6\ \sqrt{[(Interspotspacing\_x)^2 + (Interspotspacing\_y)^2]} \quad \textbf{[3]}$$

where n=1,2,3 or 4. Figure 41 shows the set of nearest neighbors where n = 2.

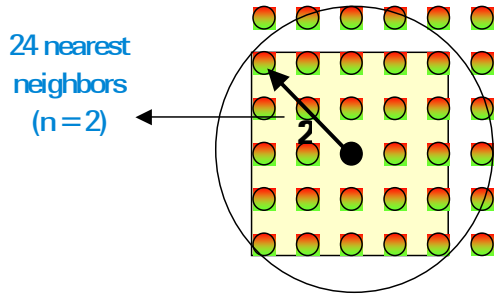


Figure 41 Example of the radius for the second closest set of nearest neighbors, or n=2

Step 5: Reject outliers

The calculation to determine the boundaries for rejection of the outlier pixels is defined below in the equations and diagram.

Assumptions for default value of 1.42 The following assumptions lead to the default value of 1.42 for this parameter.

- Normal distribution for pixel intensity, where y-axis corresponds to pixel frequency and x-axis corresponds to pixel intensity.
- A 99% confidence interval that the pixels of interest are contained within the boundaries for rejection.

The *Interquartile Range (IQR)* is the range of points under a Gaussian distribution contained between the 25th percentile mark (25% of the points are contained under the curve from the zero point to the 25th percentile mark) and the 75th percentile mark. The 50th percentile mark is coincident with the median of the curve.

The *boundary for rejection* is the point on the x-axis beyond which all pixels will be rejected.

"D" is the distance between the mean of the curve and the boundary for rejection.

Calculations of default value The following calculations are based on the above assumptions.

- If a pixel is located within the 99% confidence interval, it is 2.6 standard deviations (SD) away from the mean. Or, D = 2.6*SD and $D = Mult\_factor \times IQR + \kappa$.
- From the Z table for cumulative normal frequency distribution, the $Z_{P=0.75}$ = 0.675.

  Therefore, $\kappa = 0.675 \times$ SD = IQR/2
- If you combine the four equations above and solve for the *Mult_factor*, the *Mult_factor* = 1.42.
- If you would rather use a 95% confidence interval, IQR *Mult_factor* = 0.952. The reason for this is, assuming normal distribution and infinite degrees of freedom, D = 1.96 * SD = $0.95185 \times IQR + \kappa$.

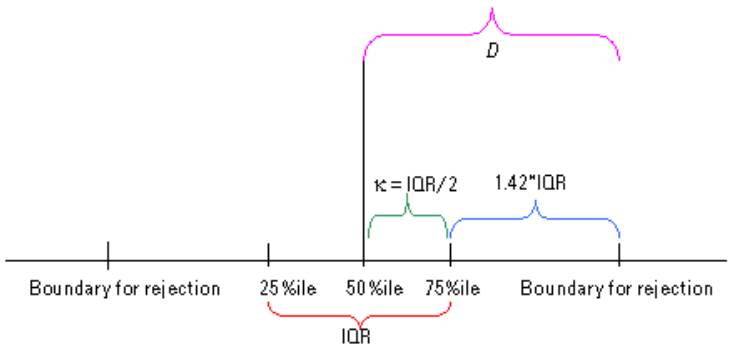


Figure 42 Important points on Gaussian curve—# of pixels vs intensity

Step 6: Calculate the mean signal of the feature (MeanSignal)

The intensities of inlier pixels of a feature are averaged to give mean signal of the feature before background subtraction. The NumPix column in the result file lists the number of inlier pixels in the cookie that remain after rejection of outlier pixels.

$$MeanSignal = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \textbf{[4]}$$

where n is the # of inlier pixels (i.e. NumPix), and X_i is pixel intensity in the feature

The number of pixels that are removed as outliers at the high end and low end of the intensity distribution are shown in 4 columns of the FEATURES table: NumPixOLLo and NumPixOLHi (for both red and green channels).

If the method in the protocol for calculating the spot value from pixel statistics has been chosen to be Median/Normalized InterQuartile Range instead of Mean/Standard Deviation , the program makes these substitutions for the spot value and background subtraction calculations:

MedianSignal for MeanSignal
BGMedianSignal for BGMean Signal
PixNorm IQR for PixSDev
GPixNormIQR for BGPixSDev
NormIQR = 0.7413 x IQR

The program does not make these substitutions for the Flag Outliers calculations.

See the previous page for the definition of the Interquartile Range (IQR).

Step 7: Calculate the mean signal of the local background (BGMeanSignal)

The intensities of local background inlier pixels are averaged to give the local background mean signal. The BGNumPix column in the result file lists the number of inlier pixels in the local background radius that remain after rejection of outlier pixels.

$$BGMeanSignal = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \textbf{[5]}$$

where n is the # of inlier pixels in the local background (i.e. BGNumPix), and X_i is the pixel intensity in the local background

Step 8: Determine if the feature is saturated (IsSaturated)

Feature is saturated if 50% of inlier pixels have intensity values above the saturation threshold.

Flag Outliers

Step 9: Determine if the feature is a non-uniformity outlier (IsFeatNonUnifOL)

σ_M^2 is the *measured variance* of inlier pixels in the feature or background (e.g. PixSDev2 or BGPixSDev2).

σ_E^2 is the *estimated variance* using known noise characteristics of the Agilent Microarray Gene Expression system.

The non-uniformity outlier algorithm flags anomalous features and local backgrounds based on statistical deviations from the Agilent noise model. Feature or background is flagged as a non-uniformity outlier (e.g. IsFeatNonUnifOL or IsBGNonUnifOL, respectively) if the *measured variance* is greater than the product of the *estimated variance* and the *confidence interval multiplier*.

$$\sigma_M^2 > (\sigma_E^2 \times CI)$$ where CI is the confidence interval calculated from chi square distribution

For more information on confidence interval, check Numerical Recipes in C (Chapter 15, page 692).

The equations below are calculated for each feature and background per channel.

Estimated Feature or Background Variance

The Agilent noise model estimates the expected variance by using noise effects from the Agilent Microarray Gene Expression system, which includes microarray manufacture, wet lab chemistry, and scanner noise.

$$\sigma^2_E = \sigma^2_{Labeling/FeatureSynthesis} + \sigma^2_{Counting} + \sigma^2_{Noise} \quad \textbf{[6]}$$

$$\sigma^2_E = Ax^2 + Bx + C \quad \textbf{[7]}$$

Net signal is the mean signal (i.e. MeanSignal or BGMeanSignal, respectively) minus the MinSigArray, which is minimum feature signal or minimum local background signal on the microarray, representing an estimate of the scanner offset.

x is the net signal of feature or background.

A or $\sigma^2_{Labeling/FeatureSynthesis}$ is the term that estimates the sources of variance that are proportional to the square of the signal, including microarray manufacturing and wet chemistry effects; the variance follows a Gaussian distribution. This term is intensity dependent and is the square of the CV (e.g. coefficient of variation) estimate of the pixel noise.

$$CV = \frac{PixSDev}{MeanSignal - MinSig_{Array}}$$

where B or $\sigma^2_{Counting}$ is the term that estimates the sources of variance that are proportional to the square-root of the signal, including scanning measurement or counting error; the variance follows a Poisson distribution. This term is dependent on the intensity and the scan resolution of the image.

where C or σ^2_{Noise} is the term that estimates the sources of variance that are independent of the signal, including electronic noise in scanner and background level noise in glass; the variance is a Constant.

The variables A, B and C have different values for feature and background. For Agilent data produced with the GE2-SSPE_95_Feb07 protocol, these values are determined empirically (default selection in protocol) from self-vs-self experiments and from the known noise characteristics of the Agilent Microarray system discussed above. For all other Agilent FE protocols, only the A term is empirically determined.

For all other Agilent protocols used with either Agilent data and GenePix data, the default selection in the protocol is to determine the B and C terms automatically. Here is how the Feature Extraction program calculates these terms:

- Saturated features are omitted from the population of negative control probes (NC). This NC set and the local background reagions associated with these features are used in the calculations.
- Calculates Net Signal.
- Calculates the pixel standard deviation and then squares it to yield the pixel variance.
- From a histogram plot of number of features or bkgd vs. net signal, finds the net signal value for the 25th percentile.
- From a histogram plot of number of feature or local bdgd vs. variance, finds the variance for the 25th percentile.

- Calculates the B term as 25%NetSignal X B Term Multiplier and the C term as 25%Variance X C Term Multiplier.

  For a given scanner, multipliers need to be determined. This tuning should use many images from different batches of microarrays, different users, and different processes. Different channels may need their own multipliers.

Measured Feature or Background Variance

$$\sigma_M^2 = \frac{1}{n-1} \times \sum_{i=0}^{n-1} (X_i - \bar{X})^2 \quad \textbf{[8]}$$

where n is # of inlier pixels in the feature or background (i.e. NumPix or BGNumPix, respectively).

where X_i is raw pixel intensity in the feature or background. (inlier pixels)

where $\bar{X}$ is mean raw pixel intensity for the feature or background (i.e. MeanSignal or BGMeanSignal, respectively).

Step 10: Determine if the feature is a population outlier (IsFeatPopOL)

Agilent provides two different statistical algorithms for identifying population outliers. You select the appropriate algorithm to use in the protocol.

For probe sequences with enough replicate features, Feature Extraction uses the IQR test for population outlier analysis. The minimum number of replicates needed is set by the protocol field, “Minimum Population” and is set to 10 as the default for most Agilent protocols.

If the protocol choice, “Use Qtest for Small Populations?” is set to True, the Q-test method is used when a probe sequence has fewer than the minimum population number of features. The Q-test choice is set to True for Agilent’s newer protocols.

Qtest for replicate features < minimum population number

Q-test allows population outlier flagging for probe sequences from one less than the minimum population number down to 3.

This test is especially useful for NegC probes on CGH microarrays. Flagging features as population outliers is needed to accurately calculate NegCAvg and SD statistics. It is also useful for the miRNA extraction where flagging features as population outliers is needed to accurately calculate Gene statistics.

This algorithm uses the following equation:

Qi = |Xi - Xnearest|\|Xmax - Xmin|

Where Xi = the intensity of a probe sequence;

Xnearest = the intensity of the nearest probe sequence in intensity

Xmax = the intensity of the most intense probe sequence

Xmin = the intensity of the least intense probe sequence

Qi is compared to Qcritical to determine if the feature is an outlier. Qcritical depends upon the number of replicate features (N) and upon the chosen confidence level.

Agilent has chosen a 95% confidence level and bases the identification of population outliers on this table:

Table 32 Qcritical values at 95% confidence level

Number of replicated features (N)	Qcritical
3	0.970
4	0.829
5	0.710
6	0.625
7	0.568

Table 32 Qcritical values at 95% confidence level

Number of replicated features (N)	Qcritical
8	0.526
9	0.493
10	0.466

IQR Test for replicate features > or = minimum population number

The equations below are calculated for each feature and background population per channel.

The intensities of all features or background regions in the population are plotted on a distribution curve. The difference in intensities between the 25th and 75th percentiles represent the Interquartile Range (IQR).

See "Step 5: Reject outliers" on page 204 for definitions to help you understand the Interquartile Range

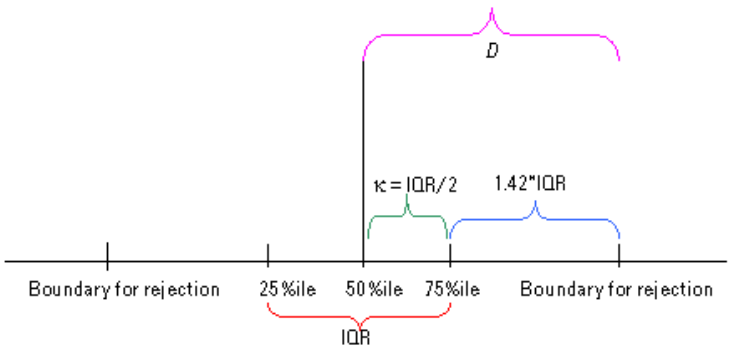


Figure 43 Interquartile Range

$$Cutoff_{PopOutlier} = 1.42 \times IQR \quad \textbf{[9]}$$

where IQR = Intensity at 75th percentile – Intensity at 25th percentile.

where 1.42 is the IQR factor. Agilent uses 1.42 as the IQR factor so that the cutoff boundaries encompass 99% of the expected population distribution. The user can change this factor to encompass different boundaries, as discussed in the Feature Extraction *User Guide*.

Feature or background is flagged as population outlier (e.g. IsFeatPopOL or IsBGPopOL, respectively) if the mean signal (e.g. MeanSignal or BGMeanSignal) is greater than the upper rejection boundary (RBupper) or less than the lower rejection boundary (RBLower).

$$MeanSignal > RB_{Upper}$$

$$MeanSignal < RB_{Lower}$$

where

$$RB_{Upper} = I_{75percentile} + Cutoff_{PopOutlier}$$

and

$$RB_{Upper} = I_{25percentile} - Cutoff_{PopOutlier}$$

Compute Bkgd, Bias and Error

Step 11: Calculate the feature background-subtracted signal (BGSubSignal)

The feature background-subtracted signal, BGSubSignal, is calculated by subtracting a value called the BGUsed from the feature mean signal.

$$BGSubSignal = MeanSignal - BGUsed \quad [10]$$

where BGSubSignal and BGUsed depend on the type of background method and the settings for spatial detrend and global background adjust. See the table below.

Table 33 Values for BGSubSignal, BGUsed and BGSDUsed for different methods and settings*

Background Subtraction Method	Background Subtraction Variable	Spatial Detrend (SpDe) OFF Global Bkgnd Adjust (GBA) OFF	SpDe ON GBA OFF	SpDe OFF GBA ON	Spatial Detrend ON Global Bkgnd Adjust ON
No background subtract	BGUsed =	BGMeanSignal†	SpatialDetrend SurfaceValue	BGAdjust	SpatialDetrendSurface Value (SDSV) + BGAdjust
	BGSDUsed =	BGPixSDev‡	BGPixSDev	BGPixSDev	BGPixSDev
	BGSubSignal =	MeanSignal	MeanSignal - BGUsed	MeanSignal - BGUsed	MeanSignal - BGUsed
Local Background	BGUsed =	BGMeanSignal	BGMeanSignal + SDSV	BGMeanSignal + BGAdjust	BGMeanSignal + SDSV + BGAdjust
	BGSDUsed =	BGPixSDev	BGPixSDev	BGPixSDev	BGPixSDev
	BGSubSignal =	MeanSignal - BGUsed	MeanSignal - BGUsed	MeanSignal - BGUsed	MeanSignal - BGUsed

Table 33 Values for BGSubSignal, BGUsed and BGSDUsed for different methods and settings*

Background Subtraction Method	Background Subtraction Variable	Spatial Detrend (SpDe) OFF Global Bkgnd Adjust (GBA) OFF	SpDe ON GBA OFF	SpDe OFF GBA ON	Spatial Detrend ON Global Bkgnd Adjust ON
Global Background method	BGUsed =	GlobalBGInlierAve** (GBGIA)	GBGIA + SDSV	GBGIA + BGAdjust	GBGIA + SDSV + BGAdjust
	BGSDUsed =	GlobalBGInlierSDev (GBGISD)	GBGISD	GBGISD	GBGISD
	BGSubSignal =	MeanSignal - BGUsed	MeanSignal - BGUsed	MeanSignal - BGUsed	MeanSignal - BGUsed

* For both the red and green channels (2-color, CGH and non-Agilent microarrays)

† With No background subtraction as the setting, BGMeanSignal is the value for BGUsed only for the t-test, but no BGUsed is subtracted from the MeanSignal to produce BGSubSignal.

‡ If the method in the protocol for calculating the spot value from pixel statistics is Median/Normalized InterQuartile Range instead of Mean/Standard Deviation, the program makes these substitutions for the spot value and background subtraction calculations:
MedianSignal for MeanSignal
BGMedianSignal for BGMeanSignal
PixNorm IQR for PixSDev
GPixNormIQR for BGPixSDev
NormIQR = 0.7413 x IQR

** If Median is the selection in the protocol, the median is substituted for the mean in the inlierAve and the InlierSDev calculations.

Step 12: Perform background spatial detrending to fit a surface

To calculate the spatial shape or surface for each channel, the Feature Extraction program uses one of these protocol selections:

- All Feature Types

  This selection fits the surface to a set of very low intensity features evenly distributed on the slide using a “moving windowed filtering”. This algorithm, which was the original algorithm for gene expression microarrays, moves a window over the whole microarray and attempts to choose a fixed number of data points with the lowest intensity inside each window.

- OnlyNegativeControlFeatures

  This selection fits the surface to the set of negative control features distributed on the slide and is recommended for Agilent CGH microarrays.

- FeaturesInNegativeControlRange

  This algorithm uses the same moving window as the first option but performs a spatial interpolation of the value of the negative controls. For interpolated negative controls, only the features that are within 3 “errors” of the fit are selected. It is recommended for Agilent GE 1 and GE 2 microarrays.

  For high density microarrays, this algorithm can take a long time to complete its calculations. To speed up the process, you can elect in the protocol to randomly select a small percentage of the total points with which to calculate the fit. To do this, you set “Perform Filtering for Fit to True, which significantly reduces the amount of time for spatial detrending of high density microarrays.

A 2D-Loess algorithm fits the surface on the mean intensities of the filtered low intensity features of both red and green channels separately. You can find more information on the algorithm from the web site, http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm.

If N = number of data points selected for surface fitting after filtering and $I_i = i^{th}$ point from the filtered low intensity data set, the Loess algorithm fits a surface through these data points to obtain an intensity value describing the surface corresponding to each input data point. Let O_i denote the fitted output surface corresponding to the i^{th} input point I_i. The statistical results that come out of this calculation are described in the table on the next page.

Table 34 Statistical results of spatial detrend algorithm

Result	Description and Equation
SpatialDetrendRMSFit	This result gives an idea of the extent of the surface fit. It is the root mean square of the fitted data points obtained from the Loess algorithm. $\sqrt{\frac{\sum_{i=1}^{N}\left(O_i - \frac{\sum_{i=1}^{N} O_i}{N}\right)^2}{N}}$
SpatialDetrendRMSFilteredminusFit	This result is the approximate residual from the surface fit. The deviations of the input (filtered) points from the corresponding output (fitted) data points are computed. An outlier rejection is performed on the set of deviations using the standard IQR technique (Figure 43 on page 211). $\sqrt{\frac{\sum_{i=1}^{N'}(I_i - O_i)^2}{N'}}$

Table 34 Statistical results of spatial detrend algorithm

Result	Description and Equation
SpatialDetrendSurfaceArea	This result gives an idea of the curvature of the surface gradient.
SpatialDetrendVolume	The volume is calculated as the sum of the intensities of the surface area minus the offset. The offset is calculated as the volume under the flat surface (parallel to the glass slide) passing through the minimum intensity point of the fitted surface. This number (total volume - offset) is normalized by the area of the microarray.
SpatialDetrendAveFit	This describes the average intensity of the surface gradient. $\frac{\sum_{i=1}^{N} O_i}{N}$

Step 13: Adjust the background

This algorithm determines the offset in both the red and green channels by identifying features that are not differentially expressed and fall within the central tendency of the data, especially in the lower intensity domain. These features should not be saturated or be flagged as non-uniform outliers. Using this method yields more accurate and reproducible background-subtracted signals and log ratios for two-channel data than using no correction or single-channel correction.

Using a self-self microarray (i.e. same target labeled in red and green channels), one expects to see a linear plot of red background-subtracted signal versus green. If the backgrounds have not been estimated correctly in one channel with respect to the second channel, there will be a bias. This bias yields a “hook” at the low end of the signal range when shown in a plot with log scale axes (see Figure 44).

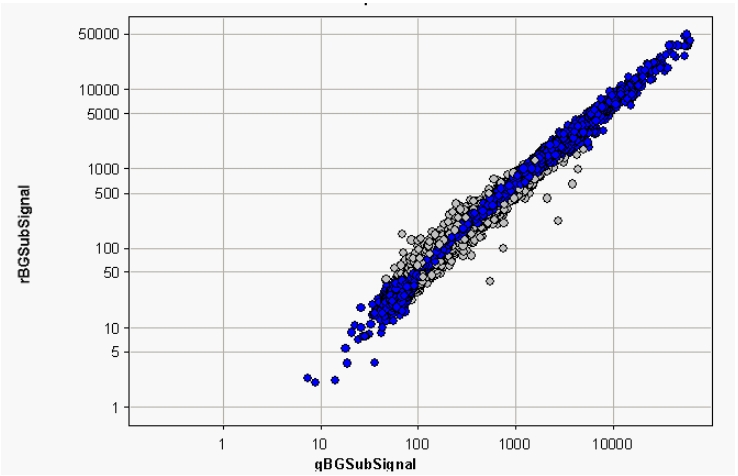


Figure 44 Unadjusted background-subtracted signals

The background adjustment algorithm first finds the central tendency of the data (features shown as blue circles in the figures). Using this subset of features, the algorithm then estimates the best adjustment in both the red and green channels to remove the bias. After the background adjustment, the bias is removed and the plot is linear (Figure 45).

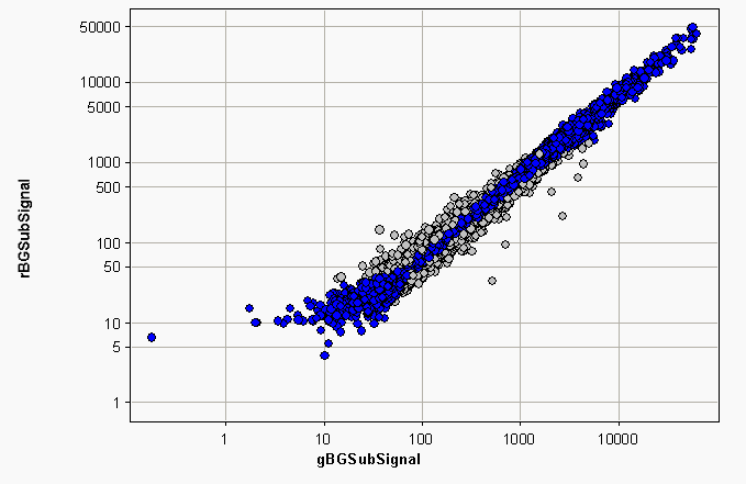


Figure 45 Adjusted background-subtracted signals

The bias, if uncorrected, yields a log ratio versus signal plot that is not symmetric about the log ratio axis (Figure 46); whereas, after adjustment, the data is more symmetric (Figure 47).

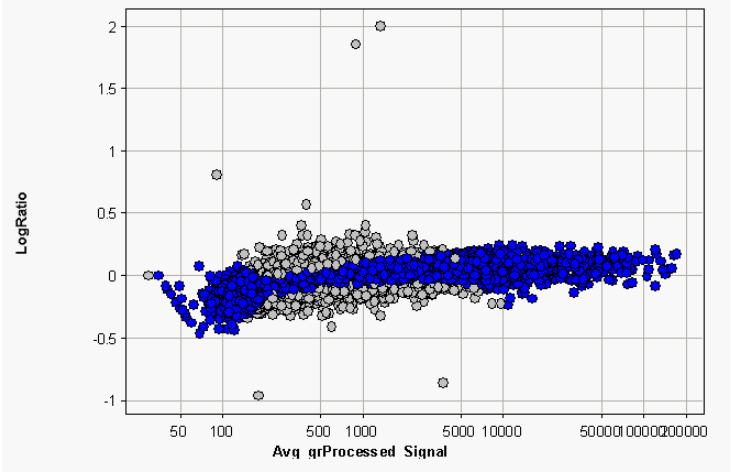


Figure 46 Log ratios calculated from unadjusted background-subtracted signals

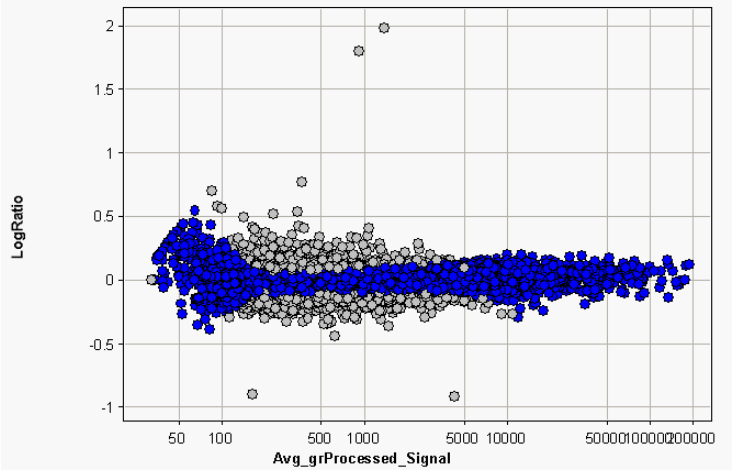


Figure 47 Log ratios calculated from adjusted background-subtracted signals

How is the Adjust background globally "pad" used? If *Adjust background globally* is selected, you can enter a constant between 0 and 500, called the pad value, which forces the log ratio of red/green towards zero. The value of the pad is expressed in raw counts, before dye normalization. The Feature Extraction program assumes that this value applies to the red or green channel with the smallest mean signal and automatically computes the corresponding raw value in the other channel that would yield a corrected log ratio of zero after dye normalization.

The red and green feature signals are analyzed for rank consistency. If red signal is plotted vs. green signal and the slope of the rank consistent features is >1, then the pad value is assigned to the green channel. If the slope is <1, the value is assigned to the red channel.

For instance, if you set *Adjust background globally to* 50, and if the slope is 1.2, then a value of 50 is added to the green background-subtracted signal of all features; whereas, a value of (50*1.2) = 60 is added to the red background-subtracted signal of all features.

Conversely, if you set *Adjust background globally to* 50, and if the slope is 0.5, then a value of 50 is added to the red background-subtracted signal of all features; whereas, a value of (50/ 0.5) = 100 is added to the green background-subtracted signal of all features.

Step 14: Calculate robust negative control statistics

This algorithm is used primarily for CGH and miRNA microarrays. It repeats the population outlier algorithm, but not on one sequence at a time, rather on the distribution of all features that are classfied as NegC or negative controls.

The algorithm calculates robust IQR statistics on features not designated as non-uniform outliers, population outliers or saturated.

UpperLimit = 75th percentile + Multiplier*IQR

LowerLimit = 25th percentile - Multiplier*IQR

The default value for this multiplier is 5.

The algorithm then omits features that are outside the Upper and LowerLimits and calculates the new robust Count, Avg, and SD of these inliers for the net signal and the background-subtracted signal:

g(r)NegCtrlNumInliers

g(r)NegCtrlAveNetSig

g(r)NegCtrlSDevNetSig

g(r)NegCtrlAveBGSubSig

g(r)NegCtrlSDevBGSubSig

Step 15: Determine the error in the signal calculation

This step calculates the error on the background-subtracted and detrended signal. You can select for the error calculation either the Universal Error Model or the model (Universal or propagated) that produces the largest (most conservative) estimate of the error.

The Feature Extraction program does a dynamic computation of an approximation for the additive terms in both the red and green channels for the Universal Error Model. The estimation of the dynamic additive error term for each channel (red or green) is based on the following equation (for 1-color gene expression, the green channel):

$$AddError = \sqrt{m_1{}^2 \sigma_{NegCtrl}{}^2 + m_2{}^2 DNF^2(RMSFit^2) + m_3{}^2 DNF^2(residual^2)} \quad \textbf{[11]}$$

where m_1 = *MultNcAutoEstimate* = 1.0

m_2 = *MultRMSAutoEstimate* = 4.0

m_3 = *MultResidualRMSAutoEstimate* = 0

DNF = LinearDyeNormFactor of the corresponding channel

Since the Additive Error is now calculated in Compute Background, Bias and Error Section, the DNF is 1 and the Variance of the NegCtrls are not scaled for the DNF either. This scaling is done to the AdditiveError after DyeNorm is completed.

$\sigma_{NegCtrl}^{2}$ = Variance of the inlier negative control

For definitions of non-uniform and population outliers, see "Change settings to flag non-uniform outliers" on page 165 of the User Guide.

where inlier negative control implies the negative controls for the corresponding channel after rejections of saturated, population and non-uniform outliers.

where *SpatialDetrendRMSFit* = RMS of the points defining the surface fit for that channel. For more details on this term, see Table 34 on page 216.

The RMSFit term drops out of the equation for microarrays of less than 5000 features.

For Agilent 8 x format oligo microarrays, the auto-estimation algorithm uses only the variance of the inlier negative controls. You can set *m1* or *m2* in equation 22 equal to zero in the protocol settings.

MultNcAutoEstimate Multiplier for the first term in the additive error equation (standard deviation of the inlier negative control). The value changes depending on the protocol used:

GE1-v1_91 and GE2-v4_91 = 0

GE2-SSPE_91 = 1.0

CGH = 1.0

MultRMSAutoEstimate Multiplier for the second term in the additive error equation (g(r)SpatialDetrendRMSFit). This term is proportional to the amount of sequence variability in the foreground.

On gene expression arrays, Agilent uses this term because there is a single sequence for all negative controls so an estimation of any sequence-dependent foreground noise using negative controls is not possible.

For CGH micrarrays, the error model choice is to make this term and *m3* zero and use only *m1* because there are a variety of sequences used for the negative controls.

GE1-v1_91 and GE2-v4_91 = 0

GE2-SSPE_91 = 4.0

CGH = 0

MultResidualRMSAutoEstimate Multiplier for the third term in the equation and is the width of the distribution of signals used in the background spatial detrending set (after the background surface has been subtracted out).

When the background detrending set includes a group of features well-distributed across the microarray with a variety of sequences, the width of the distribution of the signals of these features after background subtraction is a very good estimate of the uncertainty of the dim signals, or the additive error.

GE1-v1_91 and GE2-v4_91 = 1

GE2-SSPE_91 = 0

CGH = 0

Step 16: Calculate the significance of feature intensity relative to background (IsPosAndSignif)

The significance of the feature intensity compared to the background intensity (local or global) is calculated using two different significance tests: one using pixel statistics for both the feature and the background values and the other using the additive error from the Error Model calculation for the background value.

Significance based on pixel statistics This method to determine significance uses the 2-sided Student's t-test with mean signal for the feature and the background correction for the background. This is implemented as an incomplete Beta Function approximation.

$$t = \frac{(\overline{X_F} - \overline{X_B})}{\sqrt{\frac{(n_F - 1)\sigma_F^2 + (n_B - 1)\sigma_B^2}{df} \times \left(\frac{1}{n_F} + \frac{1}{n_B}\right)}} \quad \textbf{[12]}$$

where $\bar{X}_F$ is the mean signal (MeanSignal) of the feature and $\bar{X}_B$ is the background correction used for subtraction (BGUsed—see Table 33 on page 213).

where n_F and n_B are the number of inlier pixels in the feature or background (local), respectively (e.g. NumPix or BGNumPix).

where σ_F^2 and σ_B^2 are variances of inlier pixels for feature and background, respectively (e.g. PixSDev^2 or BGSDUsed^2).

$$\sigma_F^2 = \frac{1}{n_F - 1} \sum_{i=0}^{n-1} (X_i - \bar{X}_F)^2 \quad \textbf{[13]}$$

X_i is pixel intensity

$$\sigma_B^2 = \frac{1}{n_B - 1} \sum_{i=0}^{n-1} (X_i - \bar{X}_B)^2 \quad \textbf{[14]}$$

where *df* is the degrees of freedom,

$df = n_F + n_B - 2$

After the p-value is calculated from the 2-sided t-test using incomplete Beta Function, it is compared to the user-defined max p-value. If the calculated p-value from the Beta Function is less than the user-defined max p-value, then the feature signal is considered to be significantly different from the background signal.

If $p\text{-}value_{Calculated} < p\text{-}value_{Max}$, and if MeanSignal > BGUsed, then feature gets a Boolean flag of 1 under the IsPosAndSignif column in Feature Extraction result file.

Significance based on additive error The Error model significance also uses a Gaussian probability distribution for the calculation and tests to see if a signal is greater than 0 with a known additive error. We use the incomplete beta function, same as for Pixel significance, to compute the probability. The function uses the signals after background subtraction. That is,

instead of having a feature signal and a background signal, the test uses the feature signal and one error (second signal assumed to be zero).

The degrees of freedom are large enough to make the function Gaussian. The additive error is assumed to be 2.6 standard deviations out so AddError/2.6 is used in the tTest.

The incomplete beta function, in this case, returns the probability that the signal is 0 and compares it with the input pValue. If the probability is greater than or equal to the pValue, the feature signal is not significant. If it is less than the pValue, it is significant.

The value of the surrogate is scaled by the probability returned. The surrogate value for the Not significant signals equals AddError/2.6 * the probability, calculated this way for two reasons.

- Signals stay continuous.
- Surrogate values are not larger than the smallest significant signals.

Step 17: Determine if the feature background-subtracted signal is well above the background (IsWellAboveBG)

The feature background-subtracted signal (i.e. BGSubSignal) is compared to the noise of its background (local or global):

$$BGSubSignal > WellAboveSDMulti \times SD_{BG}$$

where

WellBoveSDMulti is the well above SD multiplier (e.g 2.6, default)

SD_{BG} is the background standard deviation (i.e. BGSDUsed)

For the Error model significance test, the SD becomes AddError/2.6.

If the background-subtracted signal is greater than the *WellAboveSDMulti* x SD_{BG}, and if the feature passes the IsPosAndSignif test, then the feature gets a Boolean flag of 1 under the IsWellAboveBG column in Feature Extraction result file.

Step 18: Calculate the surrogate value (SurrogateUsed)

The surrogate value is calculated and used as the "lowest limit of detection" to replace the dye-normalized signal when any of the following situations occur. These tests are done for each channel:

- MeanSignal is less than BGUsed or not significant compared to BGUsed (i.e., IsPosAndSignif = 0).
- BGSubSignal is less than its background standard deviation (i.e., BGSubSignal < BGSDUsed).

  The decision to replace a dye-normalized signal with a surrogate value is not made, however, until after probes are selected for correcting the dye bias.

The surrogate value is calculated in this step using these criteria:

If pixel significance is used to calculate IsPosAndSignif,

$$SurrogateUsed = SD_{BG} \quad \textbf{[15]}$$

where SD_{BG} is the background standard deviation (i.e. BGSDUsed)

If Error model significance is used to calculate IsPosAndSignif,

$$SurrogateUsed = [AddError/2.6] * p\text{-}value \quad \textbf{[16]}$$

where p-value is the Gaussian p-value and AddError is the additive error from the Error Model calculation

For the local background method, the standard deviation of the background is at the pixel-level of the local background.

For global background methods, the standard deviation of the background is at the replicate background-population level of the microarray.

Step 19: Perform multiplicative detrending

Multiplicative detrending is an algorithm designed to compensate for slight linear variations in intensities that can occur if the processing is not homogeneous across the slide. This non-homogeneous processing results in different chemical reaction times, for example, between the sides and the center, and produces a “dome effect”.

With 2-color microarrays these dome effects are the same in each channel and for the most part cancel out during the calculations. Agilent has found multiplicative detrending to still be useful, however, for all the microarrays. It is turned on in all the v.9.5 protocols, except for the GE2-nonAT_95 protocol.

This algorithm is designed to correct the data by fitting a smoothed surface via a second degree polynomial fit to the higher signals on the microarray (after outliers are rejected).

An option also exists in the 2-color gene expression protocols to detrend only on replicate signals. The algorithm normalizes replicates, fits the surface to the normalized replicates and then uses the fit to detrend the data.

Because the multiplicative trend can be confused with the additive trend for dim microarrays, data points inside a multiple times the standard deviation from the center of the signals for the negative control population are excluded. The equations for statistics and results that are produced by this calculation are shown in the following table. See Table 31, “Algorithms (Protocol Steps) and the results they produce,” on page 192 for descriptions of these results.

Table 35 Statistics and Results for Multiplicative Detrending

Results	Equation
gMultDetrendRMSFit *MDS = MultDetrendSignal*	$\sqrt{\frac{\sum_{i=1}^{N}\left(MDS_i - \frac{\sum_{i=1}^{N} MDS_i}{N}\right)^2}{N}}$
gMultDetrendSignal	$\frac{10^{Fitted(\log 10(BgSubSignal))}}{\left[\frac{\sum_{i=1}^{N}(10^{Fitted(\log 10(BgSubSignal))})_i}{N}\right]}$
gProcessedSignal	$\frac{BGSubSignal_i}{MultDetrendSignal_i}$
gProcessedSigError	$\frac{BGSubSignalError_i}{MultDetrendSignal_i}$

Correct Dye Biases

Step 20: Determine normalization features

Normalization features are features used to evaluate the dye bias between the red and green channels.

Using "All Probes" method Under this method, the initial normalization features are selected based on the following three criteria:

- Features are positive and significant versus the background (e.g. IsPosAndSignif = 1)
- Features are non-control (e.g. ControlType = 0)
- Features are non-outlier (e.g. IsFeatNonUnifOL = 0, IsFeatPopnOL = 0, IsSaturated = 0)

Using "List of Normalization Genes" method Under this method, the user selects the normalization features. These features can be housekeeping genes or genes with no differential expression.

Using "Rank Consistency Probes" method Under this method, the chosen normalization features simulate housekeeping genes. These features fall within the central tendency of the data, having consistent trends between the red and green channels. They are selected based on the following two criteria:

- Features pass the three criteria described in the "all significant, non-control, and non-outlier features" method and
- Features pass the rank consistency filter between the red and green channels

Rank consistency filter is done by transforming the feature BGSubSignal to feature rank per channel. Next, the feature correlation strength is calculated per feature:

$$CS = \frac{|\rho_R - \rho_G|}{N} \quad \textbf{[17]}$$

where ρ_R and ρ_G are the ranks of feature in the red and green channels, respectively

where N is the total number of initial normalization features

If the $CS \leq \tau$, where τ is the threshold percentile, then feature passes the rank consistency filter between the red and green channels and falls within the central tendency of the data. Note τ is a user-defined parameter in the Feature Extraction program.

Using "Rank Consistent List of Normalization Genes" This method uses the rank consistent normalization genes from the list. These genes follow the criteria described above.

Step 21: Calculate the normalization factor

LinearDyeNormFactor The linear dye normalization method assumes that dye bias is not intensity-dependent and therefore takes a global approach to dye normalization. A linear dye normalization factor is computed per channel by setting the geometric mean of signal intensity of the normalization features equal to 1000:

$$LinearDyeNormFactor = \frac{1000}{10^{\left(\frac{1}{n}\sum_{i=1}^{n}\log X_i\right)}} \quad \textbf{[18]}$$

The LinearDyeNormFactor (red and green channels) values are listed in the STATS table.

where X_i is the background-subtracted signal of a feature (i.e. BGSubSignal)

where n is the number of features used for normalization (i.e. features with IsNormalization = 1)

LOWESSDyeNormFactor The LOWESS dye normalization method assumes that dye bias may be intensity-dependent and therefore takes a local approach to dye normalization. The LOWESS dye normalization factor is calculated by fitting the locally weighted linear regression curve to the chosen normalization features. The amount of dye bias is determined from the curve at each feature's intensity. Each feature gets a different LOWESS dye normalization factor per channel.

The LOWESS method corrects the log ratio data so that its central tendency after dye normalization lies along zero for all intensity ranges, assuming an equal number of up- and down-regulated features in any given signal range. The LOWESS DyeNormFactor is derived for each channel by the procedure described on the next page:

a A linear regression curve is fit to the data in a plot of M vs A, where M (y axis) = Log(R/G) and A (x axis) = 1/2 x

Log(R*G). R and G represent the red and green background-subtracted signals. This LOWESS curve fit through the central tendency of the M vs. A plot is defined as Mfit, and is a function of A.

b The dye normalization step transforms the data so that the central tendency of Mfit at every A is shifted to be equal to zero.

c After the correction factor is determined for any feature, it is split evenly over the red and green channels.

The new signals after correction, R' and G', are obtained by transforming the original R and G:

$R' = R/(10^{MFit/2})$ and $G' = G*(10^{MFit/2})$

d If the original log ratio is exactly along the fit line Mfit, the new log ratio is shifted to zero:

If log(R/G) = Mfit, then Log(R) = Log(G) + Mfit

or $Log(R'*10^{MFit/2}) = Log(G'*10^{-MFit/2}) + Mfit$

or Log(R') + Mfit/2 = Log(G') - Mfit/2 + Mfit

or Log(R'/G') = 0

e The LOWESSDyeNormFactor for R is $1/(10^{M'/2})$.
The LOWESSDyeNormFactor for G is $10^{M'/2}$.

Note that the Linear&LOWESS dye normalization factor is not reported in the Feature Extraction output file. Therefore, the only way to know the Linear & Lowess dye norm factor is to calculate it using the equation below.

Linear&LOWESSDyeNormFactor This curve fitting algorithm does a linear scaling/normalization of the data individually in each channel before performing a non-linear dye normalization.

The Linear&LOWESS dye normalization factor can be calculated from the equation below:

$$\text{Linear\&LOWESSDyeNormFactor} = \frac{DyeNormalSignal}{BGSubSignal \times LinearDyeNormFactor} \quad \textbf{[19]}$$

Step 22: Determine if surrogate values must substitute for low-intensity signals

At this point two criteria are used to determine is surrogate values must take the place of the low-intensity signals:

- The feature signal is not positive and significant versus background.
- The signal is not larger than the background error.

Surrogate values were computed during background subtraction and are stored in the SurrogateUsed column.

Step 23: Calculate the dye-normalized signal (DyeNormSignal)

The dye-normalized signal is calculated by multiplying the background-subtracted signal by the dye normalization factor:

$$DyeNormSignal = BGSubSignal \times DNF$$

where $DNF = LinearDyeNormFactor$, when linear dye normalization method is used and where:

$$DNF = LinearDyeNormFactor \times LOWESSDyeNormFactor \quad \textbf{[20]}$$

when LOWESS dye normalization method is used.

Compute Ratios

Step 24: Calculate the processed signal (ProcessedSignal)

The processed signal is used in calculating the log ratio. If a surrogate is not used (i.e. SurrogateUsed = zero value), then the processed signal is the dye-normalized signal. If a surrogate is used (i.e. SurrogateUsed = non-zero value), then the processed signal is the SurrogateUsed value.

if $SurrogateUsed = 0$, then $ProcessedSignal = DyeNormSignal$

if $SurrogateUsed \neq 0$, then $ProcessedSignal = SurrogateUsed$

Step 25: Calculate the log ratio of feature (LogRatio)

The log ratio is the measure of differential expression between the red and green channels:

$$LogRatio = Log\left(\frac{rProcessedSignal}{gProcessedSignal}\right) \quad \textbf{[21]}$$

where *rProcessedSignal* and *gProcessedSignal* are signals post dye normalization and post surrogate processing in the red and green channels, respectively.

Step 26: Calculate the p-value and error on log ratio of feature (PvalueLogRatio and LogRatioError)

PvalueLogRatio gives the statistical significance on the log ratio per each feature (e.g. gene) between the red and green channels. The p-value is a measure of the confidence (viewed as a probability) that the feature is not differentially expressed.

For example, if the p-value is less than 0.01, we can say with a 99% confidence level that the gene is differentially expressed. In other words, there would be a 1% random chance of getting this low of a p-value with a gene that is actually not differentially expressed:

$$\text{p-value} = 1 - Erf\left(\frac{|xdev|}{\sqrt{2}}\right) = Erfc\left(\frac{|xdev|}{\sqrt{2}}\right) \quad \textbf{[22]}$$

where:

$$Erf(x) = \frac{2}{\sqrt{pi}}\int_0^x e^{-t^2}\,dt$$

Erf(x) is the error function of the expression as given by the above equation: It is twice the integral of the Gaussian distribution with mean = 0 and variance = 1/2

Erfc is the complementary error function as defined by the above equation.

xdev is the deviation of LogRatio from 0.

$$xdev = \frac{LogRatio}{LogRatioError} \quad \textbf{[23]}$$

Equation 22 is analogous to a signal to noise metric.

For more details on calculations with the Universal Error Model, see the confidential Agilent technical paper on error modeling.

If the **Universal Error Model** is used, then xdev is computed from six sources:

- ProcessedSignals (red and green channels)
- Multiplicative error factors (red and green)
- Additive error factors (red and green)

The terms xdev, 'multiplicative error', and 'additive error' come from the Universal Error Model, as developed by Rosetta Biosoftware.

Once xdev is computed, it is plugged back into Equation 2, where LogRatioError is derived.

For more details on calculations with the propagation error model, see the confidential Agilent technical paper on error modeling.

If the **Propagation of Pixel Level Error Model** is used, then LogRatioError is computed from the following sources:

- Feature PixSDev (red and green channels)
- Background Noise (calculation is dependent upon the chosen BkSubMethod; red and green channels)

Once the LogRatioError is computed, it is plugged back into Equation 21, where xdev is derived.

Calculate Metrics

Although the QC metrics are calculated in this step, only the gridding tests are discussed in this section.

Step 27: Perform a series of gridding tests to make sure that grid placement has been successful

These tests are performed to yield warnings on the Summary Reports about unsuccessful gridding. They also produce the assessment shown in the QC Report of whether the grid needs to be evaluated or not.

With FE version 9.5 new tests have been added and thresholds tuned to decrease the number of false negatives (Summary Report shows no problems when there are) and false positives (Summary Report shows a problem when there isn't).

The parameters for these tests do not appear in the protocols, but they do appear in the FEParams output.

Below is a question asked by each test, the metric used to answer the question ("stat" name that appears in the result text file as the Statistics table) and the threshold to assess gridding success or failure. If a grid fails any one of these tests, a warning or warnings appear in the reports.

Test 1 How many features are "not found" along the edge of the microarray?

Stat name: MaxSpotNotFoundEdges

Threshold_Max: 0.72

Test 2 How many local background regions are flagged as non-uniform outliers in either channel?

Stat name: AnyColorPrcntBGNonUnifOL

Threshold_Max: 2%

Test 3 How broad is the distribution of NegControl net signals?

Stat name: Max{gNegCtrlSDevNetSig, rNegCtrlSDevNetSig}

Threshold_Max: 100

Test 4 What is the median CV% of BGSubSignal of the NonControl replicated sequences?

Stat names: Max{gNegCtrlMedPrcntCVBGSubSig, rNegCtrlMedPrcntCVBGSubSig} or just the green stat for a 1-color application

Threshold_Max: 50%

Test 5 What is the difference between feature centers found by the gridding algorithm vs the spot-finding algorithm?

Stat names: Max{CentroidDiffX, CentroidDiffY}

Threshold_Max: 10%

Optional Test 6 How many features along the edge of the microarray are flagged as non-uniform outliers in either channel?

This test is used only if one of these two metrics is unavailable:

- No replicated features are present to calculate the NonCtrlMedPrcntCVBGSubSig metric.
- Or no NegControls are present to calculate the StdDev.

Stat name: MaxNonUnifEdges

Threshold_Max: 10%

MicroRNA Analysis

This step is only used for the feature extraction of microRNA microarray 1-color images.

This analysis samples multiple probes with multiple features per probe and reports the measurements and errors as the TotalGeneSignal and TotalGeneSignalError for each of the miRNAs of the 8-pack microarray. These values are reported in both the text file and a new file called the “GeneView” file.

Several steps are needed to calculate the total gene signal. First, you calculate the TotalProbeSignal and then you sum the TotalProbeSignal over the number of probes per gene.

To calculate the TotalProbeSignal and the TotalProbeError, this algorithm does the following steps:

a Calculates the EffectiveFeatureSizeFraction

b Finds the robust average of all the processd signals for each replicated probe (features with the same sequence) measured in the extraction. The same is done for the processed Signal Error column by propagating the error.

c Calculates the Nominal Spot Area S in square microns.

$$S = \pi \cdot (SpotWidth)/2 \cdot (SpotHeight)/2$$

d Multiplies each average by the total number of pixwls targeted by that probe (The total number of Features *S*EffectiveFeatureSizeFraction).

e Further multiplies by weight, where the weight is calculated as 1/30,000.

The equations and descriptions for calculating each output or result column are listed in the following table:

Table 36 Statistics and Results for the MicroRNA Analysis

Feature or Stat	Equation or Description
gTotalProbeSignal	$$\frac{\sum_{i}^{In_{Pr}} gProcSignal_{PR_i}}{In_{PR}} \cdot Tot_{PR} \cdot E \cdot S \cdot W$$ Where: PR = Index of Probe Replicates for given miRNA In = Number of replicate population inliers Tot = Total number of probe replicates E = EffectiveFeatureSizeFraction S = Nominal Spot Area - equation described on previous page W = Weight - described on previous page
gTotalProbeError	$$\sqrt{\frac{\sum_{i}^{In_{PR}} gProcSignalError_{PR_i}^{2}}{In_{PR}}} \cdot Tot_{PR} \cdot E \cdot S \cdot W$$
gTotalGeneSignal	$$\sum_{i=0}^{NumProbesPerGene} gTotalProbeSignal$$
gTotalGeneError	$$\sqrt{\sum_{i=0}^{NumProbesPerGene} gTotalProbeError^{2}}$$

Table 36 Statistics and Results for the MicroRNA Analysis

Feature or Stat	Equation or Description
IsGeneDetected	This flag considers a probe detected if the signal is three times the error. If one probe of the set of probes comprising a gene is detected, the gene is considered detected.
gEffectiveFeatureSizeFraction	Estimates the ratio of the effective feature size to the nominal feature size. It is calculated by looking at the ratio of the whole spot measurement versus the cookie measurement.
gFeatureUniformityAnaomalyFraction	Calculates the ratio of the number of features having anomalous effective feature size fractions to the total number of features. This gives a measure of the percentage of representative spots that are strange (e.g., donuts, super hot spots, or hot crescents).
gUsedDefaultEffectiveFeatureSize	Reports whether an effective feature size was estimated or not. Stat value is 0 if Yes and 1 if No. If No, the default effective feature size value is used.

Example calculations for feature 4738 of Agilent Yeast_2x11K image

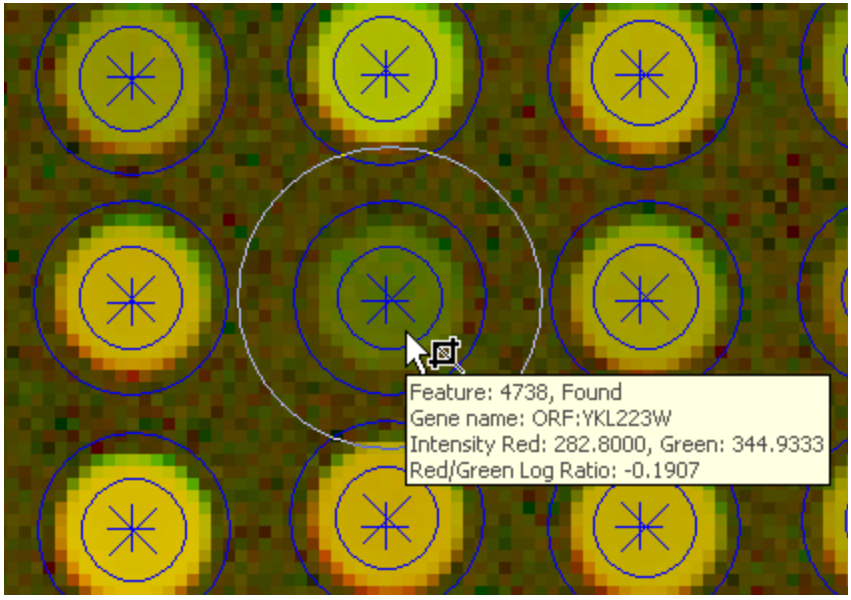


Figure 48 Figure 1 Visual results of feature number 4738 from "Shapes" file (*.shp file)

The 11k microarray image, "Yeast_2x11K", is included in the Feature Extraction program installation CD.

Data from the FEPARAMS table

BGSubtractor_BGSubMethod	BGSubtractor_BackgroundCorrectionOn	BGSubtractor_SpatialDetrendOn
7	0	1

The BGSubMethod of 7 corresponds to No Background Subtraction method (see Table 20 on page 113 of this guide.). Global Background Adjustment is turned Off. Spatial Detrending is turned On.

Data from the STATS Table

gLinearDyeNormFactor	rLinearDyeNormFactor
4.00048	1.98389

Data from the FEATURES Table

Results from Find And Measure Spots Algorithm

FeatureNum	gNumPix	rNumPix	gMeanSignal	rMeanSignal	gPixSDev	rPixSDev
4738	60	60	344.9333	282.8	6.954737	8.79869

Results from Correct Bkgd and Signal Biases Algorithm

FeatureNum	gSpatialDetrendSurfaceValue	rSpatialDetrendSurfaceValue
4738	335.81	276.468

FeatureNum	gBGUsed	rBGUsed	gBGSDUsed	rBGSDUsed	gBGSubSignal	rBGSubSignal
4738	335.81	276.468	4.44485	6.87425	9.12361	6.3325

FeatureNum	gIsPosAndSignif	rIsPosAndSignif	gIsWellAboveBG	rIsWellAboveBG
4738	1	1	0	0

rBGUsed=rSpatialDetrendSurfaceValue

276.468 = 276.468

Note that this equation is valid only if there is no background subtraction, spatial detrending is on, and there is no global background adjustment.

rBGUsed=MeanSignal - rGBGUsed

6.3325 = 282.8 – 276.468

For an explanation of BGUsed with other background settings, see Table 33 on page 213.

Results from Correct Dye Biases Algorithm

FeatureNum	gDyeNormSignal	rDyeNormSignal
4738	27.7894	16.5003

rBGUsed=rBGSubSignal x rLinearDyeNormFactor x rLOWESSDyeNormFactor

16.5003 = 6.3325 x 1.98389 x 1.313

Results from Compute Ratios and Errors Algorithm

FeatureNum	gSurrogateUsed	rSurrogateUsed	gProcessedSignal	rProcessedSignal
4738	0	17.9119	27.78941	17.91191

FeatureNum	LogRatio	LogRatioError	PValueLogRatio
4738	-0.1907373	1.021414518	0.851865755

For the red channel, does the feature number 4738 pass the two criteria listed below that are required to calculate an accurate and reproducible log ratio?

- Feature is positive and significant vs background (i.e. IsPosAndSignif = 1.
- BGSubSignal is greater than its background standard deviation (i.e. BGSDUsed).

While feature number 4738 passed criteria 1, it failed the 2nd criteria since its rBGSubSignal is LESS than rBGSDUsed.

As a result, the surrogate value is calculated as 1 standard deviation of the background multiplied by the dye normalization factors in the red channel.

rSurrogateUsed = *rBGSDUsed x rLinearDyeNormFactor* x *rLOWESSDyeNormFactor*

17.9119 = 6.87425 x 1.9838 x 1.313

Since a surrogate is used in the red channel, the red processed signal uses the surrogate value.

rProcessedSignal = *rSurrogateUsed*, if *rSurrogateUsed* ≠ 0

17.9119 = 17.9119

The Log ratio is the log of red processed signal over green processed signal.

$$LogRatio = \log \frac{rProcessedSignal}{gProcessedSignal}$$

-0.190737 = log (17.9119 / 27.78941)

It is important to note that log ratio and p-value calculations are computed differently, depending on whether a surrogate is used in only one channel, both channels, or neither channels.

Since feature 4738 uses a surrogate in only the red channel (Case 2 below) and the red surrogate value is not greater than the green processed signal, the p-value and error on the log ratio are calculated, as usual, using equations 1 and 2 in "Step 26: Calculate the p-value and error on log ratio of feature (PvalueLogRatio and LogRatioError)" on page 233 of this guide.

Case 1: R/G Both channels use DyeNormSignals. P-value and log ratio are calculated as usual. For signals not using surrogates, the g(r)DyeNormsignal is equal to the g(r)ProcessedSignal, used to calculate log ratios.	Case 2: r/G r = rSurrogateUsed G = gDyeNormSignal P-value and log ratio are calculated as usual. If r/G > 1, then FE software automatically sets LogRatio = 0 and pValueLogRatio = 1.
Case 3: R/g R = DyeNormSignal g = gSurrogateUsed P-value and log ratio are calculated as usual. If R/g < 1, then FE software automatically sets LogRatio = 0 and pValueLogRatio = 1.	Case 4: r/g Both channels use surrogates. FE software automatically sets LogRatio = 0 and pValueLogRatio = 1. For signals using surrogates, the g(r)Processed Signal is equal to the g(r)SurrogateUsed value,

Figure 49 Summary—Use of surrogates for calculations

6
Command Line Feature Extraction

The command line version of Feature Extraction (FE) software is called *FeNoWindows*. You can run FeNoWindows from any directory. The Feature Extraction installation includes FeNoWindows along with the necessary grid templates and protocols. The installer places **FeNoWindows.exe** in the **Feature Extraction** directory, and edits the **System Path Variable** to include the **Feature Extraction** directory.

NOTE When you start FeNoWindows, you cannot return to Feature Extraction until FeNoWindows completes any running tasks and exits (or exits due to an error).

FeNoWindows accepts v.8.5, v.9.1 and v.9.5 project files as input for running Feature Extraction. An FE8 project file is an XML file that specifies an extraction set. You create project files using the Feature Extraction user interface.

FeNoWindows returns result information in XML format; the result looks similar to a project XML file. FeNoWindows appends a result code to the project XML file that indicates the basic status of the run, such as successful completion, unsuccessful attempts, warnings, or errors. For a complete listing of return codes, see Table 37 on page 250.

Commands

Command line syntax

```
FeNoWindows  [-c command] [-t number]  [-o output_file ] [-p
   protocol] [-q <linktype]  <input_file> [<input_file>] …
```

command can be any of the following:

```
[extract | addgrid | addprotocol | adddyenormlist |
   removegrid | removeprotocol | removedyenormlist |
   linkprotocoltogrid | exportprotocols |
   exportdyenormlists]
```

If you do not specify a command, it defaults to `extract`.

Commands and arguments

extract This command runs Feature Extraction on the input project.

```
FeNoWindows -c extract [-t number]   [-o <output_file> ]
   <input_file>
```

input_file The name of an xml project file with the extension **.fep**.

output_file The name of the result .xml file. This file looks like a project file with the status added (as described below).

number Specifies the number of threads to use. This value can be `1` (the default), `2`, or `4`.

CAUTION You must specify the -o option when specifying the output file name, or FeNoWindows will not create the file.

NOTE FeNoWindows accepts no more than one project file as input.

addgrid This command adds a grid to the local database.

```
FeNoWindows –c addgrid [ <design_file_path> |
   <grid_file_path>]
```

design_file_path The path and name of a design file.

grid_file_path The path and name of a grid file.

addprotocol This command adds a protocol to the database.

```
FeNoWindows –c addprotocol[<protocol_file_path>]
```

protocol_file_path The path and name of a protocol file.

addmetricset This command adds a metric set to the database.

```
FeNoWindows –c addmetricset[<metricset_file_path>]
```

metricset_file_path The path and name of a metric set file.

adddyenormlist This command adds a dyenormlist to the database.

```
FeNoWindows –c adddyenormlist [-g gridtemplatename]
   <dyenormlist_file_path>
```

gridtemplatename The name of the database grid template that the probes in the dye norm list must match

dyenormlist_file_path The path and name of the dye norm list

The dye norm list needs to look like:

ProbeName1 GeneName1 SystematicName1
ProbeName2 GeneName2 SystematicName2
ProbeName3 GeneName3 SystematicName3

Spaces between words must be a tab, and no white space is allowed at the end of the file. When a list is read into the database, it is checked against the specified grid template to

make sure that the probes match with what is in the grid template. The basename of the file is used to name the dye norm list in the database.

Example:

```
-c adddyenormlist -g 14850_D_F_20060807 C:\
DyeNormlist\MyNormlist.txt
```

removegrid This command removes a grid from the database.

```
FeNoWindows -c removegrid <gridname>
```

gridname The name of the grid.

removeprotocol This command removes a protocol from the database.

```
FeNoWindows -c removeprotocol <protocol_name>
```

protocol_name The path to the protocol file.

removemetricset This command removes a metric set from the database.

```
FeNoWindows -c removemetricset <metricset_name>
```

metricset_name The path to the metric set file.

removedyenormlist This command removes a dyenormlist from the database.

```
FeNoWindows -c removedyenormlist [-g gridtemplatename]
  <dyenormlistname>
```

gridtemplatename Name of the grid template associated with the dye norm list to be removed

dyenormlistname Name of the dye norm list to be removed

Example:

```
FeNoWindows -c removedyenormlist -g
14850_D_F_20060807 MyNormlist
```

linkprotocoltogrid This command links a protocol to a grid template so that the protocol is automatically assigned if a valid scan barcode exists.

Command example: `FeNoWindows -c linkprotocoltogrid`

```
   -p myOneColorProtocol
   -q OneColor 012345_D_
   20050212

FeNoWindows -c linkprotocoltogrid [-p protocol] [-q linktype
   <gridname> ]
```

linktype Type of link, either OneColor or TwoColor, that links protocol to grid template

exportprotocols This command exports all the protocols in a given database to the location you specify.

```
FeNoWindows -c exportprotocols <to_directory>
```

to_directory The complete path to the directory where you want to keep the protocols.

exportmetricsets This command exports all the metric sets in a given database to the location you specify.

```
FeNoWindows -c exportmetricsets <to_directory>
```

to_directory The complete path to the directory where you want to keep the metric sets.

exportdyenormlists This command exports all the dyenormlists in a given database to the location you specify.

```
FeNoWindows -c exportdyenormlists <to_directory>
```

to_directory The complete path to the directory where you want to keep the dye norm lists.

Example:

```
FeNoWindows -c exportdyenormlists C:\DyeNormList
```

Return Codes

Return codes are integers that represent errors that caused FeNoWindows to fail without generating output.

They are listed in Table 37.

Table 37 FeNoWindows return codes

Return code	Description
0	The extraction project completed without errors. The output file contains extraction information for every extraction. This success code does not guarantee the validity of every extraction in the set.
1	The input parameter was not found. Check that the filename and path are correct, or that the database entry exists and is spelled correctly.
2	Invalid input file. Check that you specified a valid input file name.
3	Request ignored. If you receive this code when you are adding a protocol or grid template, the object already exists in the database and will not be added. If you receive this code when you are deleting objects, the object was not found in the database.
4	No license, or invalid license. Check the existence, location, and expiration date of your Feature Extraction license.
5	Initialization failure – MFC failed to initialize. Call tech support.
6	Initialization failure – COM failed to initialize. Call tech support.
7	Invalid command line arguments. Check spelling and syntax.

Return code	Description
8	Feature extraction failed. Call tech support.
9	Feature Extraction failed to add or remove a protocol. Database could be down. Restart the database by rebooting or starting the AGTFEDB service from the control panel.
10	Feature Extraction failed to add or remove a grid template. Restart the database.
11	The grid template or protocol link failed. Restart the database.

Extraction Input

The input file for extraction is an FE project (standard, not on-time) file with a filetype of XML.

An example of a project file (.fep) is provided below. To create project files, use the Feature Extraction user interface and the instructions in the *Quick Start Guide*.

Project Properties Settings

Note that `MAGEOutPkgType` *and* `TextOutPkgType` *are Full. This means all the features are sent to the output file. A compact subset of features is the alternate choice.*

See Chapter 3 and Chapter 4 of the Reference Guide for a listing of the FULL and COMPACT sets of features sent to the text and MAGE-ML result files.

```
<FeatureExtractionML>
   <FEPMLVerInfo VerMaj="2" VerMin="0"/>
   <FEProject Operator="Unknown"
              ResultsDirectory="C:\GridComparison\
              9.5.1.1"
              ResultsLocationSameAsImage="False"
              OutputMAGE="False"
              MAGEOutPkgType="Full"
              OutputMAGECompressed="False"
              OutputJPEG="False"
              OutputText="True"
              TextOutPkgType="Full"
              OutputVisualResults="True"
              OutputGRID="False"
              OutputArrayQCReport="True"
              FTPSendTiffFile="False"
              FTPMachineDestination=""
              FTPPort="21"
              FTPUserName="resolverftp" FTPPassword=""
              FTPProfileDestinationFolder="mage"
              OverWritePreviousResults="False"
              RDAUserName=""
              RDACtrlGroups=""
```

```
        DefaultQCMetricSet=""
        ExternalDyeNormList=""
        DefaultProtocol=""
        UseGridFileIfAvailable="False"
        UseProjDefProtocolFirst="False">
```

Example of XDR extraction set

```
    <Extraction Name="US45102874_251494710148_S01">
        <XDRScanID Name="01122007125846"/>
        <Image Name="C:\GridComparison\
        US45102874_251494710148_S01_H.tif"/>
        <ImageXDR2
        Name="US45102874_251494710148_S01_L.tif
        "/>
        <Grid Name="014947_D_20060807"
        IsGridFile="False"/>
        <Protocol Name="miRNA_95_16Jan"/>
    </Extraction>
```

Example of regular extraction set

```
    <Extraction Name="US14702375_251494710059_S01">
        <Image Name="C:\GridComparison\
        US14702375_251494710059_S01.tif"/>
        <Grid Name="014947_D_20060807"
        IsGridFile="False"/>
        <Protocol Name="miRNA_95_16Jan"/>
    </Extraction>
```

Example of extraction set with grid file

```
    <Extraction Name="US14702375_251494710059_S01">
        <Image Name="C:\GridComparison\
        US14702375_251494710059_S01.tif"/>
        <Grid Name=" C:\GridComparison\
        gridfile_grid.csv" IsGridFile="True"/>
        <Protocol Name="miRNA_95_16Jan"/>
    </Extraction>
  </FEProject>
</FeatureExtractionML>
```

Extraction Results

The information contained in the **output file (**specified with the -o command) depends on the extraction operation performed and the options you specified. For example, the XML file can contain status, time, warning or error messages, and indicate the number of outliers. Status information (Success, Error, Warning) is particularly important.

Status information

Success Feature Extraction had no issues extracting the data.

Warning Feature Extraction generated the data, which might be usable. Users should check the RTF file for the warning. Feature Extraction probably ran OK. A common warning is “No SpikeIns found on this design.”

Error Output files may or may not have been generated. If output files were generated, users need to look at the image and shape files to make sure they are OK. The grid may not have been placed correctly. Users should not trust the data without visual inspection.

FeNoWindows occasionally reports failures that are not true errors. The image, RTF file and QC report, and possibly the shapes file, need to be examined to see why things failed.

Examples of status information

The following XML file fragments show you examples of what the status information might look like (presented in red) after an extraction set is run.

Each of these messages is associated with an extraction set that has been run.

```
<Result Status="Warning" >

<ResultMessages Status=”Warning” Message="148
(Green) saturated features" />

…

<Result Status="Success" >

<ResultMessages Status=”Info” Message="10 (Green)
NonUniform Outliers" />
```

Index

P

Q

R

S

T

U

www.agilent.com

In This Book

The *Reference Guide* presents descriptions of the protocols, or methods, available for use with the Agilent Feature Extraction software (v9.5), as well as a listing of results and an explanation of how the FE algorithms work.

This guide provides:

- a list of the default settings for each protocol shipped with the software
- a list of all the parameters and results available after feature extraction
- the equations and a sample calculation for the feature extraction process.

Printed in USA
Fourth Edition, February 2007

Agilent Technologies